Linear Statistical Inference and its Applications

Second Editon

C. RADHAKRISHNA RAO
Pennsylvania State University

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.

This Page Intentionally Left Blank

Linear Statistical Inference and its Applications

A complete list of the titles in this series appears at the end of this volume.

Linear Statistical Inference and its Applications

Second Editon

C. RADHAKRISHNA RAO
Pennsylvania State University

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.

This text is printed on acid-free paper. ♾

Paperback edition published 2002.

Published simultaneously in Canada.

For ordering and customer service, call 1-800-CALL-WILEY.

Library of Congress Cataloging in Publication Data is available.

ISBN 0-471-21875-8

Printed in the United States of America

10 9 8 7 6 5 4 3

To Bhargavi

This Page Intentionally Left Blank

Preface

The purpose of this book is to present up-to-date theory and techniques of statistical inference in a logically integrated and practical form. Essentially, it incorporates the important developments in the subject that have taken place in the last three decades. It is written for readers with a background knowledge of mathematics and statistics at the undergraduate level.

Quantitative inference, if it were to retain its scientific character, could not be divested of its logical, mathematical, and probabilistic aspects. The main approach to statistical inference is inductive reasoning, by which we arrive at "statements of uncertainty." The rigorous expression that degrees of uncertainty require are furnished by the mathematical methods and probability concepts which form the foundations of modern statistical theory. It was my awareness that advanced mathematical methods and probability theory are indispensable accompaniments in a self-contained treatment of statistical inference that prompted me to devote the first chapter of this book to a detailed discussion of vector spaces and matrix methods and the second chapter to a measure-theoretic exposition of probability and development of probability tools and techniques.

Statistical inference techniques, if not applied to the real world, will lose their import and appear to be deductive exercises. Furthermore, it is my belief that in a statistical course emphasis should be given to both mathematical theory of statistics and to the application of the theory to practical problems. A detailed discussion on the application of a statistical technique facilitates better understanding of the theory behind the technique. To this end, in this book, live examples have been interwoven with mathematical results. In addition, a large number of problems are given at the end of each chapter. Some are intended to complement main results derived in the body of the chapter, whereas others are meant to serve as exercises for the reader to test his understanding of theory.

The selection and presentation of material to cover the wide field of statistical inference have not been easy. I have been guided by my own experience in teaching undergraduate and graduate students, and in conducting and guiding research in statistics during the last twenty years. I have selected and presented the essential tools of statistics and discussed in detail their theoretical bases to enable the readers to equip themselves for consultation work or for pursuing specialized studies and research in statistics.

Why Chapter 1 provides a rather lengthy treatment of the algebra of vectors and matrices needs some explanation. First, the mathematical treatment of statistical techniques in subsequent chapters depends heavily on vector spaces and matrix methods; and second, vector and matrix algebra constitute a branch of mathematics widely used in modern treatises on natural, biological, and social sciences. The subject matter of the chapter is given a logical and rigorous treatment and is developed gradually to an advanced level. All the important theorems and derived results are presented in a form readily adaptable for use by research workers in different branches of science.

Chapter 2 contains a systematic development of the probability tools and techniques needed for dealing with statistical inference. Starting with the axioms of probability, the chapter proceeds to formulate the concepts of a random variable, distribution function, and conditional expectation and distributions. These are followed by a study of characteristic functions, probability distributions in infinite dimensional product spaces, and all the important limit theorems. Chapter 2 also provides numerous propositions, which find frequent use in some of the other chapters and also serve as good equipment for those who want to specialize in advanced probability theory.

Chapter 3 deals with continuous probability models and the sampling distributions needed for statistical inference. Some of the important distributions frequently used in practice, such as the normal, Gamma, Cauchy, and other distributions, are introduced through appropriate probability models on physical mechanisms generating the observations. A special feature of this chapter is a discussion of problems in statistical mechanics relating to the equilibrium distribution of particles.

Chapter 4 is devoted to inference through the technique of analysis of variance. The Gauss-Markoff linear model and the associated problems of estimation and testing are treated in their wide generality. The problem of variance-components is considered as a special case of the more general problem of estimating intraclass correlation coefficients. A unified treatment is provided of multiclassified data under different sampling schemes for classes within categories.

The different theories and methods of estimation form the subject matter of Chapter 5. Some of the controversies on the topic of estimation are examined; and to remove some of the existing inconsistencies, certain modifications are introduced in the criteria of estimation in large samples.

Problems of specification, and associated tests of homogeneity of parallel samples and estimates, are dealt with in Chapter 6. The choice of a mathematical model from which the observations could be deemed to have arisen is of fundamental importance because subsequent statistical computations will be made on the framework of the chosen model. Appropriate tests have been developed to check the adequacy of proposed models on the basis of available facts.

Chapter 7 provides the theoretical background for the different aspects of statistical inference, such as testing of hypotheses, interval estimation, experimentation, the problem of identification, nonparametric inference, and so on.

Chapter 8, the last chapter, is concerned with inference from multivariate data. A special feature of this chapter is a study of the multivariate normal distribution through a simple characterization, instead of through the density function. The characterization simplifies the multivariate theory and enables suitable generalizations to be made from the univariate theory without further analysis. It also provides the necessary background for studying multivariate normal distributions in more general situations, such as distributions on Hilbert space.

Certain notations have been used throughout the book to indicate sections and other references. The following examples will help in their interpretation. A subsection such as **4f·3** means subsection 3 in section f of Chapter 4. Equation (4f.3.6) is the equation numbered 6 in subsection **4f·3** and Table 4f.3β is the table numbered second in subsection **4f·3**. The main propositions (or theorems) in each subsection are numbered: (i), (ii), etc. A back reference such as [(iii). **5d·2**] indicates proposition (iii) in subsection **5d·2**.

A substantial part of the book was written while I was a visiting professor at the Johns Hopkins University, Baltimore, in 1963–1964, under a Senior Scientist Fellowship scheme of the National Science Foundation, U.S.A. At the Johns Hopkins University, I had the constant advice of G. S. Watson, Professor of Statistics, who read the manuscript at the various stages of its preparation. Comments by Herman Chernoff on Chapters 7 and 8, by Rupert Miller and S. W. Dharmadhikari on Chapter 2, and by Ralph Bradley on Chapters 1 and 3, have been extremely helpful in the preparation of the final manuscript. I wish to express my thanks to all of them. The preparation and revision of the manuscript would not have been an easy task without the help of G. M. Das, who undertook the heavy burden of typing and organizing the manuscript for the press with great care and diligence.

Finally, I wish to express my gratitude to the late Sir Ronald A. Fisher and to Professor P. C. Mahalanobis under whose influence I have come to appreciate statistics as the new technology of the present century.

Calcutta, India C. R. Rao
June, 1965

This Page Intentionally Left Blank

Preface to the Second Edition

As in the first edition, the aim has been to provide in a single volume a full discussion of the wide range of statistical methods useful for consulting statisticians and, at same time, to present in a rigorous manner the mathematical and logical tools employed in deriving statistical procedures, with which a research worker should be familiar.

A good deal of new material is added, and the book is brought up to date in several respects, both in theory and applications.

Some of the important additions are different types of generalized inverses, concepts of statistics and subfields, MINQUE theory of variance components, the law of iterated logarithms and sequential tests with power one, analysis of dispersion with structural parameters, discrimination between composite hypotheses, growth models, theorems on characteristic functions, etc.

Special mention may be made of the new material on estimation of parameters in a linear model when the observations have a *possibly singular covariance matrix*. The existing theories and methods due to Gauss (1809) and Aitken (1935) are applicable only when the covariance matrix is known to be nonsingular. The new *unified approaches* discussed in the book (Section 4i) are valid for all situations whether the covariance matrix is singular or not.

A large number of new exercises and complements have been added.

I wish to thank Dr. M. S. Avadhani, Dr. J. K. Ghosh, Dr. A. Maitra, Dr. P. E. Nüesch, Dr. Y. R. Sarma, Dr. H. Toutenberg, and Dr. E. J. Williams for their suggestions while preparing the second edition.

New Delhi C. R. Rao

This Page Intentionally Left Blank

Contents

CHAPTER 1

Algebra of Vectors and Matrices 1

1a. Vector Spaces 2

1a.1 Definition of Vector Spaces and Subspaces, 1a.2 Basis of a Vector Space, 1a.3 Linear Equations, 1a.4 Vector Spaces with an Inner Product

Complements and Problems 11

1b. Theory of Matrices and Determinants 14

1b.1 Matrix Operations, 1b.2 Elementary Matrices and Diagonal Reduction of a Matrix, 1b.3 Determinants, 1b.4 Transformations 1b.5 Generalized Inverse of a Matrix, 1b.6 Matrix Representation, of Vector Spaces, Bases, etc., 1b.7 Idempotent Matrices, 1b.8 Special Products of Matrices

Complements and Problems 30

1c. Eigenvalues and Reduction of Matrices 34

1c.1 Classification and Transformation of Quadratic Forms, 1c.2 Roots of Determinantal Equations, 1c.3 Canonical Reduction of Matrices, 1c.4 Projection Operator, 1c.5 Further Results on g-Inverse, 1c.6 Restricted Eigenvalue Problem

1d. Convex Sets in Vector Spaces 51

1d.1 Definitions, 1d.2 Separation Theorems for Convex Sets

1e. Inequalities 53

1e.1 Cauchy-Schwarz (C-S) Inequality, 1e.2 Hölder's Inequality, 1e.3 Hadamard's Inequality, 1e.4 Inequalities Involving Moments, 1e.5 Convex Functions and Jensen's Inequality, 1e.6 Inequalities in Information Theory, 1e.7 Stirling's Approximation

1f. Extrema of Quadratic Forms 60

1f.1 General Results, 1f.2 Results Involving Eigenvalues and Vectors 1f.3 Minimum Trace Problems

Complements and Problems 67

CHAPTER 2
Probability Theory, Tools and Techniques 79

2a. Calculus of Probability 80

2a.1 The Space of Elementary Events, 2a.2 The Class of Subsets (Events), 2a.3 Probability as a Set Function, 2a.4 Borel Field (σ-field) and Extension of Probability Measure, 2a.5 Notion of a Random Variable and Distribution Function, 2a.6 Multidimensional Random Variable, 2a.7 Conditional Probability and Statistical Independence, 2a.8 Conditional Distribution of a Random Variable

2b. Mathematical Expectation and Moments of Random Variables 92

2b.1 Properties of Mathematical Expectation, 2b.2 Moments, 2b.3 Conditional Expectation, 2b.4 Characteristic Function (c.f.), 2b.5 Inversion Theorems, 2b.6 Multivariate Moments

2c. Limit Theorems 108

2c.1 Kolmogorov Consistency Theorem, 2c.2 Convergence of a Sequence of Random Variables, 2c.3 Law of Large Numbers, 2c.4 Convergence of a Sequence of Distribution Functions, 2c.5 Central Limit Theorems, 2c.6 Sums of Independent Random Variables

2d. Family of Probability Measures and Problems of Statistics 130

2d.1 Family of Probability Measures, 2d.2 The Concept of a Sufficient Statistic, 2d.3 Characterization of Sufficiency

Appendix 2A Stieltjes and Lebesgue Integrals 132

Appendix 2B. Some Important Theorems in Measure Theory and Integration 134

Appendix 2C. Invariance 138

Appendix 2D. Statistics, Subfields, and Sufficiency 139

Appendix 2E. Non-Negative Definiteness of a Characteristic Function 141

Complements and Problems 142

CHAPTER 3
Continuous Probability Models 155

3a. Univariate Models 158

3a.1 Normal Distribution, 3a.2 Gamma Distribution, 3a.3 Beta Distribution, 3a.4 Cauchy Distribution, 3a.5 Student's t Distribution, 3a.6 Distributions Describing Equilibrium States in Statistical Mechanics, 3a.7 Distribution on a Circle

3b. Sampling Distributions 179

3b.1 Definitions and Results, 3b.2 Sum of Squares of Normal Variables, 3b.3 Joint Distribution of the Sample Mean and Variance, 3b.4 Distribution of Quadratic Forms, 3b.5 Three Fundamental Theorems of the Least Squares Theory, 3b.6 The p-Variate Normal Distribution, 3b.7 The Exponential Family of Distributions

3c. Symmetric Normal Distribution 197

3c.1 Definition, 3c.2 Sampling Distributions

3d. Bivariate Normal Distribution 201

3d.1 General Properties, 3d.2 Sampling Distributions

Complements and Problems 209

CHAPTER 4
The Theory of Least Squares and Analysis of Variance 220

4a. Theory of Least Squares (Linear Estimation) 221

4a.1 Gauss-Markoff Setup (Y, $X\beta$, $\sigma^2 I$), 4a.2 Normal Equations and Least Squares (l.s.) Estimators, 4a.3 g-Inverse and a Solution of the Normal Equation, 4a.4 Variances and Covariances of l.s. Estimators, 4a.5 Estimation of σ^2, 4a.6 Other Approaches to the l.s. Theory (Geometric Solution), 4a.7 Explicit Expressions for Correlated Observations, 4a.8 Some Computational Aspects of the l.s. Theory, 4a.9 Least Squares Estimation with Restrictions on Parameters, *4a.10 Simultaneous Estimation of Parametric Functions, 4a.11 Least Squares Theory when the Parameters Are Random Variables, 4a.12 Choice of the Design Matrix*

4b. Tests of Hypotheses and Interval Estimation 236

4b.1 Single Parametric Function (Inference), 4b.2 More than One Parametric Function (Inference), 4b.3 Setup with Restrictions

4c. Problems of a Single Sample 243

4c.1 The Test Criterion, 4c.2 Asymmetry of Right and Left Femora (Paired Comparison)

4d. One-Way Classified Data 244

4d.1 The Test Criterion, 4d.2 An Example

4e. Two-Way Classified Data 247

4e.1 Single Observation in Each Cell, 4e.2 Multiple but Equal Numbers in Each Cell, 4e.3 Unequal Numbers in Cells

4f. A General Model for Two-Way Data and Variance Components 258

4f.1 A General Model, 4f.2 Variance Components Model, 4f.3 Treatment of the General Model

4g. The Theory and Application of Statistical Regression 263

4g.1 Concept of Regression (General Theory), 4g.2 Measurement of Additional Association, 4g.3 Prediction of Cranial Capacity (a Practical Example), 4g.4 Test for Equality of the Regression Equations, 4g.5 The Test for an Assigned Regression Function, 4g.6 Restricted Regression

4h. The General Problem of Least Squares with Two Sets of Parameters 288

4h.1 Concomitant Variables, 4h.2 Analysis of Covariance, 4h.3 An Illustrative Example

4i. Unified Theory of Linear Estimation 294

4i.1 A Basic Lemma on Generalized Inverse, 4i.2 The General Gauss-Markoff Model (GGM), 4i.3 The Inverse Partitioned Matrix (IPM) Method, 4i.4 Unified Theory of Least Squares

4j. Estimation of Variance Components 302

4j.1 Variance Components Model, 4j.2 MINQUE Theory, 4j.3 Computation under the Euclidian Norm

4k. Biased Estimation in Linear Models 305

4k.1 Best Linear Estimator (BLE), 4k.2 Best Linear Minimum Bias Estimation (BLIMBE)

Complements and Problems 308

CHAPTER 5

Criteria and Methods of Estimation 314

5a. Minimum Variance Unbiased Estimation 315

5a.1 Minimum Variance Criterion, 5a.2 Some Fundamental Results on Minimum Variance Estimation, 5a.3 The Case of Several Parameters, 5a.4 Fisher's Information Measure, 5a.5 An Improvement of Unbiased Estimators

5b. General Procedures 334

5b.1 Statement of the General Problem (Bayes Theorem), 5b.2 Joint d.f. of (θ, x) Completely Known, 5b.3 The Law of Equal Ignorance, 5b.4 Empirical Bayes Estimation Procedures, 5b.5 Fiducial Probability, 5b.6 Minimax Principle, 5b.7 Principle of Invariance

5c. Criteria of Estimation in Large Samples 344

5c.1, Consistency, 5c.2 Efficiency

5d. Some Methods of Estimation in Large Samples 351

5d.1 Method of Moments, 5d.2 Minimum Chi-Square and Associated Methods, 5d.3 Maximum Likelihood

5e. Estimation of the Multinomial Distribution 355

5e.1 Nonparametric Case, 5e.2 Parametric Case

5f. Estimation of Parameters in the General Case 363

5f.1 Assumptions and Notations, 5f.2 Properties of m.l. Equation Estimators

5g. The Method of Scoring for the Estimation of Parameters 366

Complements and Problems 374

CHAPTER 6

Large Sample Theory and Methods 382

6a. Some Basic Results 382

6a.1 Asymptotic Distribution of Quadratic Functions of Frequencies, 6a.2 Some Convergence Theorems

6b. Chi-Square Tests for the Multinomial Distribution 390

6b.1 Test of Departure from a Simple Hypothesis, 6b.2 Chi-Square Test for Goodness of Fit, 6b.3 Test for Deviation in a Single Cell, 6b.4 Test Whether the Parameters Lie in a Subset, 6b.5 Some Examples, 6b.6 Test for Deviations in a Number of Cells

6c. Tests Relating to Independent Samples from Multinomial Distributions 398

6c.1 General Results, 6c.2 Test of Homogeneity of Parallel Samples, 6c.3 An Example

6d. Contingency Tables 403

6d.1 The Probability of an Observed Configuration and Tests in Large Samples, 6d.2 Tests of Independence in a Contingency Table, 6d.3 Tests of Independence in Small Samples

6e. Some General Classes of Large Sample Tests 415

6e.1 Notations and Basic Results, 6e.2 Test of a Simple Hypothesis, 6e.3 Test of a Composite Hypothesis

6f. Order Statistics 420

6f.1 The Empirical Distribution Function, 6f.2 Asymptotic Distribution of Sample Fractiles

6g. Transformation of Statistics 426

6g.1 A General Formula, 6g.2 Square Root Transformation of the Poisson Variate, 6g.3 Sin^{-1} *Transformation of the Square Root of the Binomial Proportion, 6g.4* Tanh^{-1} *Transformation of the Correlation Coefficient*

6h. Standard Errors of Moments and Related Statistics 436

6h.1 Variances and Covariances of Raw Moments, 6h.2 Asymptotic Variances and Covariances of Central Moments, 6h.3 Exact Expressions for Variances and Covariances of Central Moments

Complements and Problems 439

CHAPTER 7

Theory of Statistical Inference 444

7a. Testing of Statistical Hypotheses 445

7a.1 Statement of the Problem, 7a.2 Neyman-Pearson Fundamental Lemma and Generalizations, 7a.3 Simple H_0 *against Simple* H, *7a.4 Locally Most Powerful Tests, 7a.5 Testing a Composite Hypothesis, 7a.6 Fisher-Behrens Problem, 7a.7 Asymptotic Efficiency of Tests*

7b. Confidence Intervals 470

7b.1 The General Problem, 7b.2 A General Method of Constructing a Confidence Set, 7b.3 Set Estimators for Functions of θ

7c. Sequential Analysis 474

7c.1 Wald's Sequential Probability Ratio Test, 7c.2 Some Properties of the S.P.R.T., 7c.3 Efficiency of the S.P.R.T., 7c.4 An Example of Economy of Sequential Testing, 7c.5 The Fundamental Identity of Sequential Analysis, 7c.6 Sequential Estimation, 7c.7 Sequential Tests with Power One

7d. Problem of Identification—Decision Theory 491

7d.1 Statement of the Problem, 7d.2 Randomized and Nonrandomized Decision Rules, 7d.3 Bayes Solution, 7d.4 Complete Class of Decision Rules, 7d.5 Minimax Rule

7e. Nonparametric Inference 497

7e.1 Concept of Robustness, 7e.2 Distribution-Free Methods, 7e.3 Some Nonparametric Tests, 7e.4 Principle of Randomization

7f. Ancillary Information 505

Complements and Problems 506

CHAPTER 8

Multivariate Analysis 516

8a. Multivariate Normal Distribution 517

8a.1 Definition, 8a.2 Properties of the Distribution, 8a.3 Some Characterizations of N_p, 8a.4 Density Function of the Multivariate Normal Distribution, 8a.5 Estimation of Parameters, 8a.6 N_p as a Distribution with Maximum Entropy

8b. Wishart Distribution 533

8b.1 Definition and Notation, 8b.2 Some Results on Wishart Distribution

8c. Analysis of Dispersion 543

8c.1 The Gauss-Markoff Setup for Multiple Measurements, 8c.2 Estimation of Parameters, 8c.3 Tests of Linear Hypotheses, Analysis of

Dispersion (A.D.), 8c.4 Test for Additional Information, 8c.5 The Distribution of Λ, 8c.6 Test for Dimensionality (Structural Relationship), 8c.7 Analysis of Dispersion with Structural Parameters (Growth Model)

8d. Some Applications of Multivariate Tests 562

8d.1 Test for Assigned Mean Values, 8d.2 Test for a Given Structure of Mean Values, 8d.3 Test for Differences between Mean Values of Two Populations, 8d.4 Test for Differences in Mean Values between Several Populations, 8d.5 Barnard's Problem of Secular Variations in Skull Characters

8e. Discriminatory Analysis (Identification) 574

8e.1 Discriminant Scores for Decision, 8e.2 Discriminant Analysis in Research Work, 8e.3 Discrimination between Composite Hypotheses

8f. Relation between Sets of Variates 582

8f.1 Canonical Correlations, 8f.2 Properties of Canonical Variables, 8f.3 Effective Number of Common Factors, 8f.4 Factor Analysis

8g. Orthonormal Basis of a Random Variable 587

8g.1 The Gram-Schmidt Basis, 8g.2 Principal Component Analysis

Complements and Problems 593

Publications of the Author 605

Author Index 615

Subject Index 618

Linear Statistical Inference and Its Applications

This Page Intentionally Left Blank

Chapter 1

ALGEBRA OF VECTORS AND MATRICES

Introduction. The use of matrix theory is now widespread in both pure mathematics and the physical and the social sciences. The theory of vector spaces and transformations (of which matrices are a special case) have not, however, found a prominent place, although they are more fundamental and offer a better understanding of problems. The vector space concepts are essential in the discussion of topics such as the theory of games, economic behavior, prediction in time series, and the modern treatment of univariate and multivariate statistical methods.

The aim of the first chapter is to introduce the reader to the concepts of vector spaces and the basic results. All important theorems are discussed in great detail to enable the beginner to work through the chapter. Numerous illustrations and problems for solution are given as an aid to further understanding of the subject. To introduce wide generality (this is important and should not cause any difficulty in understanding the theory) the elements used in the operations with vectors are considered as belonging to any *Field* in which *addition* and *multiplication* are defined in a consistent way (as in the ordinary number system). Thus, the elements $e_1, e_2, \ldots$ (finite or infinite in number) are said to belong to a field F, if they are closed under the operations of addition $(e_i + e_j)$ and multiplication $(e_i e_j)$, that is, sums and products of elements of F also belong to F, and satisfy the following conditions:

(A_1) $e_i + e_j = e_j + e_i$ (commutative law)

(A_2) $e_i + (e_j + e_k) = (e_i + e_j) + e_k$ (associative law)

(A_3) For any two elements e_i, e_j, there exists an element e_k such that $e_i + e_k = e_j$.

The condition (A_3) implies that there exists an element e_0 such that $e_i + e_0 = e_i$ for all i. The element e_0 is like 0 (zero) of the number system.

(M_1) $e_i e_j = e_j e_i$ (commutative law)

(M_2) $e_i(e_j e_k) = (e_i e_j)e_k$ (associative law)

(M_3) $e_i(e_j + e_k) = e_i e_j + e_i e_k$ (distributive law)

(M_4) For any two elements e_i and e_j such that $e_i \neq e_0$, the zero element, there exists an element e_k such that $e_i e_k = e_j$.

(M_4) implies that there is an element e_1 such that $e_i e_1 = e_i$ for all i. The element e_1 is like 1 (unity) of the number system.

The study of vector spaces is followed by a discussion of the modern matrix theory and quadratic forms. Besides the basic propositions, a number of results used in mathematical physics, economics, biology and statistics, and numerical computations are brought together and presented in a unified way. This would be useful for those interested in applications of the matrix theory in the physical and the social sciences.

1a VECTOR SPACES

1a.1 Definition of Vector Spaces and Subspaces

Concepts such as force, size of an organism, an individual's health or mental abilities, and price level of commodities cannot be fully represented by a single number. They have to be understood by their manifestations in different directions, each of which may be expressed by a single number. The mental abilities of an individual may be judged by his scores $(x_1, x_2, \ldots, x_k)$ in k specified tests. Such an ordered set of measurements may be simply represented by $\mathbf{x}$, called a row vector. If $\mathbf{y} = (y_1, y_2, \ldots, y_k)$ is the vector of scores for another individual, the total scores for the two individuals in the various tests may be represented by $\mathbf{x} + \mathbf{y}$ with the definition

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_k + y_k). \tag{1a.1.1}$$

This rule of combining or adding two vectors is the same as that for obtaining the resultant of two forces in two or three dimensions, known as the *parallelogram law of forces*. Algebraically this law is equivalent to finding a force whose components are the sum of the corresponding components of the individual forces.

Given a force $\mathbf{f} = (f_1, f_2, f_3)$, it is natural to define

$$c\mathbf{f} = (cf_1, cf_2, cf_3) \tag{1a.1.2}$$

as a force c times the first, which introduces a new operation of multiplying a vector by a scalar number such as c. Further, given a force $\mathbf{f}$, we can counterbalance it by adding a force $\mathbf{g} = (-f_1, -f_2, -f_3) = (-1)\mathbf{f}$, or by applying $\mathbf{f}$ in the opposite direction, we have the resulting force $\mathbf{f} + \mathbf{g} = (0, 0, 0)$. Thus we have the concepts of a negative vector such as $-\mathbf{f}$ and a null vector $\mathbf{0} = (0, 0, 0)$, the latter having the property $\mathbf{x} + \mathbf{0} \doteq \mathbf{x}$ for all $\mathbf{x}$.

It is seen that we are able to work with the new quantities, to the extent permitted by the operations defined, in the same way as we do with numbers. As a matter of fact, we need not restrict our algebra to the particular type of new quantities, *viz.*, ordered sets of scalars, but consider a collection of elements $\mathbf{x}, \mathbf{y}, \mathbf{z}, \ldots$, finite or infinite, which we choose to call vectors and $c_1, c_2, \ldots$, scalars constituting a *field* (like the ordinary numbers with the operations of addition, subtraction, multiplication, and division suitably defined) and lay down certain rules of combining them.

Vector Addition. The operation of addition indicated by + is defined for any two vectors leading to a vector in the set and is subject to the following rules:

(i)	$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$	(commutative law)
(ii)	$\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$	(associative law)

Null Element. There exists an element in the set denoted by $\mathbf{0}$ such that

(iii) $$\mathbf{x} + \mathbf{0} = \mathbf{x}, \qquad \text{for all} \quad \mathbf{x}.$$

Inverse (Negative) Element. For any given element $\mathbf{x}$, there exists a corresponding element $\boldsymbol{\xi}$ such that

(iv) $$\mathbf{x} + \boldsymbol{\xi} = \mathbf{0}.$$

Scalar Multiplication. The multiplication of a vector $\mathbf{x}$ by a scalar c leads to a vector in the set, represented by $c\mathbf{x}$, and is subject to the following rules.

(v)	$c(\mathbf{x} + \mathbf{y}) = c\mathbf{x} + c\mathbf{y}$	(distributive law for vectors)
(vi)	$(c_1 + c_2)\mathbf{x} = c_1\mathbf{x} + c_2\mathbf{x}$	(distributive law for scalars)
(vii)	$c_1(c_2\mathbf{x}) = (c_1c_2)\mathbf{x}$	(associative law)
(viii)	$e\mathbf{x} = \mathbf{x}$	(where e is the unit element in the field of scalars).

A collection of elements (with the associated field of scalars F) satisfying the axioms (i) to (viii) is called a *linear vector space* $\mathscr{V}$ or more explicitly $\mathscr{V}(F)$ or $\mathscr{V}_F$. Note that the conditions (iii) and (iv) can be combined into the single condition that for any two elements $\mathbf{x}$ and $\mathbf{y}$ there is a unique element $\mathbf{z}$ such that $\mathbf{x} + \mathbf{z} = \mathbf{y}$.

The reader may satisfy himself that, for a given k, the collection of all ordered sets $(x_1, \ldots, x_k)$ of real numbers with the addition and the scalar multiplication as defined in (1a.1.1) and (1a.1.2) form a vector space. The same is true of all polynomials of degree not greater than k, with coefficients belonging to any field and addition and multiplication by a constant being defined in the usual way. Although we will be mainly concerned with vectors which are ordered sets of real numbers in the treatment of statistical methods covered in this book, the axiomatic set up will give the reader proper insight into the new algebra and also prepare the ground for a study of more complicated vector spaces, like Hilbert space (see Halmos, 1951), which are being increasingly used in the study of advanced statistical methods.

A *linear subspace, subspace,* or *linear manifold* in a vector space $\mathscr{V}$ is any subset of vectors $\mathscr{M}$ closed under addition and scalar multiplication, that is, if $\mathbf{x}$ and $\mathbf{y} \in \mathscr{M}$, then $(c\mathbf{x} + d\mathbf{y}) \in \mathscr{M}$ for any pair of scalars c and d. Any such subset $\mathscr{M}$ is *itself* a vector space with respect to the same definition of addition and scalar multiplication as in $\mathscr{V}$. The subset containing the null vector alone, as well as that consisting of all the elements in $\mathscr{V}$, are extreme examples of subspaces. They are called improper subspaces whereas others are proper subspaces.

As an example, all linear combinations of a given fixed set S of vectors $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k$ is a subspace called the linear manifold $\mathscr{M}(S)$ spanned by S. This is the smallest subspace containing S.

Consider k linear equations in n variables $x_1, \ldots, x_n$,

$$a_{i1} x_1 + \cdots + a_{in} x_n = 0, \qquad i = 1, \ldots, k,$$

where a_{ij} belongs to any field F. The reader may verify that the totality of solutions $(x_1, \ldots, x_n)$ considered as vectors constitutes a subspace with the addition and scalar multiplication as defined in (1a.1.1) and (1a.1.2).

1a.2 Basis of a Vector Space

A set of vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ is said to be linearly dependent if there exist scalars $c_1, \ldots, c_k$, not all simultaneously zero, such that $c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k = \mathbf{0}$, otherwise it is independent. With such a definition the following are true:

1. The null vector by itself is a dependent set.
2. Any set of vectors containing the null vector is a dependent set.
3. A set of non-zero vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ is dependent when and only when a member in the set is a linear combination of its predecessors.

A linearly independent subset of vectors in a vector space $\mathscr{V}$, generating or spanning $\mathscr{V}$ is called a basis (*Hamel* basis) of $\mathscr{V}$.

(i) *Every vector space $\mathscr{V}$ has a basis.*

To demonstrate this, let us choose sequentially non-null vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots,$ in $\mathscr{V}$ such that no $\boldsymbol{\alpha}_i$ is dependent on its predecessors. In this process it may so happen that after the kth stage no independent vector is left in $\mathscr{V}$, in which case $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k$ constitute a basis and $\mathscr{V}$ is said to be a finite (k) dimensional vector space. On the other hand, there may be no limit to the process of choosing $\boldsymbol{\alpha}_i$, in which case $\mathscr{V}$ is said to have infinite dimensions. Further argument is needed to show the actual existence of an infinite set of independent vectors which generate all the vectors. This is omitted as our field of study will be limited to finite dimensional vector spaces. The following results concerning finite dimensional spaces are important.

(ii) *If* $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k$ *and* $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s$ *are two alternative choices for a basis, then* $s = k$.

Let, if possible, $s > k$. Consider the dependent set $\boldsymbol{\beta}_1, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k$. If $\boldsymbol{\alpha}_i$ depends on the predecessors, then $\mathscr{V}$ can also be generated by $\boldsymbol{\beta}_1, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{i-1}, \boldsymbol{\alpha}_{i+1}, \ldots, \boldsymbol{\alpha}_k$, in which case the set $\boldsymbol{\beta}_2, \boldsymbol{\beta}_1, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{i-1}, \boldsymbol{\alpha}_{i+1}, \ldots, \boldsymbol{\alpha}_k$ is dependent. One more $\boldsymbol{\alpha}$ can now be omitted. The process of adding a $\boldsymbol{\beta}$ and omitting an $\boldsymbol{\alpha}$ can be continued (observe that no $\boldsymbol{\beta}$ can be eliminated at any stage) till the set $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k$ is left, which itself spans $\mathscr{V}$ and hence $(s - k)$ of the $\boldsymbol{\beta}$ vectors are redundant. The *cardinal number k* common to all bases represents the minimal number of vectors needed to span the space or the maximum number of independent vectors in the space. We call this number as the dimension or rank of $\mathscr{V}$ and denote it by $d[\mathscr{V}]$.

(iii) *Every vector in* $\mathscr{V}$ *has a unique representation in terms of a given basis.*

If $\sum a_i \boldsymbol{\alpha}_i$ and $\sum b_i \boldsymbol{\alpha}_i$ represent the same vector, then $\sum (a_i - b_i)\boldsymbol{\alpha}_i = \mathbf{0}$, which is not possible unless $a_i - b_i = 0$ for all i, since $\boldsymbol{\alpha}_i$ are independent.

The Euclidean space $E_k(F)$ of all ordered sets of k elements $(x_1, \ldots, x_k)$, $x_i \in F$ (a field of elements) is of special interest. The vectors may be considered as points of k dimensional Euclidean space. The vectors $\mathbf{e}_1 = (1, 0, \ldots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \ldots, 0), \ldots, \mathbf{e}_k = (0, 0, \ldots, 1)$ are in E_k and are independent, and any vector $\mathbf{x} = (x_1, \ldots, x_k) = x_1\mathbf{e}_1 + \cdots + x_k\mathbf{e}_k$. Therefore $d[E_k] = k$, and the vectors $\mathbf{e}_1, \ldots, \mathbf{e}_k$ constitute a basis and thus any other independent set of k vectors. Any vector in E_k can, therefore, be represented as a unique linear combination of k independent vectors, and, naturally, as a combination (not necessarily unique) of *any set* of vectors containing k independent vectors.

When F is the field of *real numbers* the vector space $E_k(F)$ is denoted simply by R^k. In the study of design of experiments we use *Galois fields* (GF) with a finite number of elements, and consequently the vector space has only a finite number of vectors. The notation $E_k(GF)$ may be used for such spaces. The vector space $E_k(F)$ when F is the field of complex numbers is represented by U^k and is called a k dimensional unitary space. The treatment of Section **1a.3** is valid for any F. Later we shall confine our attention to R^k only. We prove an important result in (iv) which shows that study of finite dimensional vector spaces is equivalent to that of E_k.

(iv) *Any vector space* $\mathscr{V}_F$ *for which* $d[\mathscr{V}] = k$ *is isomorphic to* E_k.

If $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k$ is a basis of $\mathscr{V}_F$, then an element $\mathbf{u} \in \mathscr{V}_F$ has the representation $\mathbf{u} = a_1\boldsymbol{\alpha}_1 + \cdots + a_k\boldsymbol{\alpha}_k$ where $a_i \in F$, $i = 1, \ldots, k$. The correspondence

$$\mathbf{u} \to \mathbf{u}^* = (a_1, \ldots, a_k), \qquad \mathbf{u}^* \in E_k(F)$$

establishes the isomorphism. For if $\mathbf{u} \to \mathbf{u}^*$, $\mathbf{v} \to \mathbf{v}^*$, then $\mathbf{u} + \mathbf{v} \to \mathbf{u}^* + \mathbf{v}^*$ and $c\mathbf{u} \to c\mathbf{u}^*$. This result also shows that any two vector spaces with the same dimensions are isomorphic.

1a.3 Linear Equations

Let $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m$ be *m fixed* vectors in an arbitrary vector space $\mathscr{V}_F$ and consider the linear equation in the scalars $x_1, \ldots, x_m \in F$ (associated Field)

$$x_1\boldsymbol{\alpha}_1 + \cdots + x_m\boldsymbol{\alpha}_m = \mathbf{0}. \tag{1a.3.1}$$

(i) *A necessary and sufficient condition that* (1a.3.1) *has a nontrivial solution, that is, not all x_i simultaneously zero, is that $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m$ should be dependent.*

(ii) *The solutions considered as (row) vectors $\mathbf{x} = (x_1, \ldots, x_m)$ in $E_m(F)$ constitute a vector space.*

This is true for if $\mathbf{x}$ and $\mathbf{y}$ are solutions then $a\mathbf{x} + b\mathbf{y}$ is also a solution. Note that $\boldsymbol{\alpha}_i$, themselves may belong to any vector space $\mathscr{V}_F$.

(iii) *Let $\mathscr{S}$ be the linear manifold or the subspace of solutions and $\mathscr{M}$, that of the vectors $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m$. Then $d[\mathscr{S}] = m - d[\mathscr{M}]$ where the symbol d denotes the dimensions of the space.*

Without loss of generality let $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k$ be independent, that is, $d[\mathscr{M}] = k$, in which case,

$$\boldsymbol{\alpha}_j = a_{j1}\boldsymbol{\alpha}_1 + \cdots + a_{jk}\boldsymbol{\alpha}_k, \qquad j = k + 1, \ldots, m.$$

We observe that the vectors

$$\begin{aligned} \boldsymbol{\beta}_1 &= (a_{k+1,1}, \ldots, a_{k+1,k}, -1, 0, \ldots, 0) \\ &\cdots \qquad \cdots \qquad \cdots \\ \boldsymbol{\beta}_{m-k} &= (a_{m,1}, \ldots, a_{m,k}, 0, 0, \ldots, -1) \end{aligned} \tag{1a.3.2}$$

are independent and satisfy the equation (1a.3.1). The set (1a.3.2) will be a basis of the solution space $\mathscr{S}$ if it spans all the solutions. Let $\mathbf{y} = (y_1, \ldots, y_m)$ be any solution and consider the vector

$$\mathbf{y} + y_{k+1}\boldsymbol{\beta}_1 + \cdots + y_m\boldsymbol{\beta}_{m-k}, \tag{1a.3.3}$$

which is also a solution. But this is of the form $(z_1, \ldots, z_k, 0, \ldots, 0)$ which means $z_1\boldsymbol{\alpha}_1 + \cdots + z_k\boldsymbol{\alpha}_k = 0$. This is possible only when $z_1 = \cdots = z_k = 0$.