

# APPLIED LINEAR STATISTICAL MODELS

---

Second Edition

John Neter  
William Wasserman  
Michael H. Kutner

---

# **Applied Linear Statistical Models**

Regression, Analysis of Variance,  
and Experimental Designs

**John Neter**  
University of Georgia

**William Wasserman**  
Syracuse University

**Michael H. Kutner**  
Emory University

**1985      Second edition**

***IRWIN***

Homewood, Illinois 60430

---

## Preface

---

Linear statistical models for regression, analysis of variance, and experimental designs are widely used today in business administration, economics, and the social, health, and biological sciences. Successful applications of these models require a sound understanding of both the underlying theory and the practical problems that are encountered in using the models in real-life situations. While *Applied Linear Statistical Models*, Second Edition, is basically an applied book, it seeks to blend theory and applications effectively, avoiding the extremes of presenting theory in isolation and of giving elements of applications without the needed understanding of the theoretical foundations.

The second edition differs from the first in a number of important respects.

1. We have added some important new topics. In the area of regression models, we have introduced discussions of detection of multicollinearity, ridge regression, and detection of influential observations. In recent years, noteworthy new developments in the detection of multicollinearity and influential observations have taken place, and a current text in linear models needs to cover these topics adequately. We have also added a chapter on nonlinear regression to introduce the reader to nonlinear statistical models.

In the area of experimental designs, we have added two chapters on nested designs, including rules for finding expected mean squares and sums of squares for a wide variety of balanced nested and crossed designs.

We have also added some specialized topics, such as  $P$ -values of test statistics, the method of least absolute deviations, analysis of variance techniques when the treatment means have unequal importance, and single degree of freedom tests.

2. We have reorganized and expanded a number of topics. In the area of

regression analysis, these include weighted regression, selection of independent variables, and normal probability plots. In analysis of variance, we have greatly expanded the discussion of the unbalanced case, devoting a full chapter to the two-factor unbalanced case including a consideration of missing cells. In the discussion of experimental designs, we have unified the treatment of missing observations in terms of the regression approach, and have split the material on randomized block designs into two chapters to facilitate understanding. At the same time we have made extensive revisions of other materials on the basis of classroom experience to improve the clarity of the presentation.

3. We have strengthened the integration of regression analysis, analysis of variance, and experimental designs through the general linear model, and have simplified the discussion by considering the same coding of indicator variables for all analysis of variance models.

4. We have placed greater emphasis on graphic methods for diagnosis, analysis, and presentation of results, and have introduced a number of computer-generated plots to demonstrate the usefulness of computer graphics.

5. The scope of the examples has been expanded to include applications from the health and biological sciences, in addition to applications from management, economics, and the social sciences. In all cases, an application can be readily understood by the general reader, regardless of background.

6. We have added three extensive real-world data sets that can be employed in a variety of ways.

7. Finally, we have substantially expanded the problem materials at the ends of the chapters and have grouped them into three categories, namely *Problems*, *Exercises*, and *Projects*. The *Problems* category includes basic problems and questions, the *Exercises* category includes conceptual and theoretical questions, and the *Projects* category includes problems utilizing large data sets and/or involving extensive calculations and analysis.

The first 15 chapters of the Second Edition of *Applied Linear Statistical Models* have also been published as a separate book under the title *Applied Linear Regression Models*.

A key feature of *Applied Linear Statistical Models* is its unified approach to the application of linear statistical models in regression, analysis of variance, and experimental designs. Instead of treating these areas in isolated fashion, we seek to show the interrelationships between them. Use of a common notation for regression on the one hand and analysis of variance and experimental designs on the other facilitates a unified view. The notion of a general linear statistical model, which arises naturally in the context of regression models, is carried over to analysis of variance and experimental design models to bring out their relation to regression models. This unified approach also has the advantage of simplified presentation.

We have included in this book not only the more conventional topics in regression, analysis of variance, and basic experimental designs, but also have taken up topics that are frequently slighted though important in practice. Thus, we devote a full chapter to indicator variables, covering both dependent and inde-

pendent indicator variables. Another chapter takes up computer-assisted selection procedures for obtaining a “best” set of independent variables to be employed in the regression model. The use of residual analysis for examining the aptness of the model is a recurring theme throughout this book. So is the use of remedial measures that may be helpful when the model is not appropriate. In the analysis of the results of a study, we emphasize the use of estimation procedures, rather than tests, because estimation is often more meaningful in practice. Also, since practical problems seldom are concerned with a single comparison, we stress the use of multiple comparison procedures.

Theoretical ideas are presented to the degree needed for good understanding in making sound applications. Proofs are given in those instances where we feel they serve to demonstrate an important method of approach. Emphasis is placed on a thorough understanding of the models, particularly the meaning of the model parameters, since such understanding is basic to proper applications. A wide variety of case examples is presented to illustrate the use of the theoretical principles, to show the great diversity of applications of linear statistical models, and to demonstrate how analyses are carried out for different problems.

We use “Notes” and “Comments” sections in each chapter to present additional discussion and matters related to the mainstream of development. In this way, the basic ideas in a chapter are presented concisely and without distraction. Similarly, optional “Topics” chapters supplement chapters containing the main development and present a variety of additional topics that in most cases can be omitted without loss of continuity.

Applications of linear statistical models frequently require extensive computations. We take the position that a computer is available in most applied work. Further, almost every computer user has access to program packages for regression analysis and analysis of variance of different types. Hence, we explain the basic mathematical steps in fitting a linear statistical model but do not dwell on computational details. This approach permits us to avoid many complex formulas and enables us to focus on basic principles. We make extensive use in this text of computer capabilities for performing computations and illustrate a variety of computer printouts and explain how these are used for analysis.

A selection of problems is provided at the end of each chapter (excepting Chapter 1). Here the reader can reinforce his or her understanding of the methodology and use the concepts learned to analyze data. We have been careful to supply data-analysis problems that typify genuine applications. In most problems the calculations are best handled on a calculator or computer, and we urge that this avenue be used when possible.

We assume that the reader of *Applied Linear Statistical Models* has had an introductory course in statistical inference, covering the material outlined in Chapter 1. Should some gaps in the reader’s background exist, he or she can read the relevant portions of an introductory text, or the instructor of the class may use supplemental materials for covering the missing segments. Chapter 1 is primarily intended as a reference chapter of basic statistical results for continuing use as the reader progresses through the book.

Calculus is not required for reading *Applied Linear Statistical Models*. In a number of instances we use calculus to demonstrate how some important results are obtained, but these demonstrations are confined to supplementary comments or notes and can be omitted without any loss of continuity. Readers who do know calculus will find these comments and notes in natural sequence so that the benefits of the mathematical developments are obtained in their immediate context. Some basic elements of matrix algebra are needed for linear models, in general, and for multiple regression, in particular. Chapter 6 introduces these elements of matrix algebra in the context of simple regression for easy learning.

*Applied Linear Statistical Models* is intended for use in undergraduate or graduate courses in linear statistical models and in second courses in applied statistics. The extent to which material presented in this text is used in a particular course depends upon the amount of time available and the objectives of the course. Some possible courses include:

1. A two-quarter or two-semester course in regression, analysis of variance, and basic experimental designs might be based on the following chapters:  
     Regression: 2, 3, 4, 5 (Sections 5.1–5.4), 6 (Sections 6.1–6.12), 7, 8, 10 (Sections 10.1–10.3), 12.  
     Analysis of variance: 16, 17, 18, 20, 21, 25.  
     Experimental designs: 26, 27, 28, 31.
2. A one-quarter or one-semester course in regression analysis might be based on the following chapters:  
     2, 3, 4, 5 (Sections 5.1–5.4), 6 (Sections 6.1–6.12), 7, 8, 9, 10 (Sections 10.1–10.3), 11 (selected topics), 12, 13, 14.
3. A one-quarter or one-semester course in analysis of variance might be based on the following chapters:  
     16, 17, 18, 19 (selected topics), 20, 21, 22, 23 (selected topics), 24, 25.
4. A one-quarter or one-semester course in regression and analysis of variance might be based on the following chapters:  
     Regression: 2, 3, 4, 5 (Sections 5.1–5.4), 6 (Sections 6.1–6.12), 7, 8, 10 (Sections 10.1–10.3).  
     Analysis of variance: 16, 17, 18, 20, 21.
5. A one-quarter or one-semester course in basic experimental designs might be based on the following chapters:  
     26, 27, 28, 29, 30, 31.

As time permits, the instructor could cover additional topics in the text.

This book can also be used for self-study by persons engaged in the fields of business administration, economics, and the social, health, and biological sciences who desire to obtain competence in the application of linear statistical models.

A book such as this cannot be written without substantial assistance from others. We are indebted to the many contributors who have developed the theory and practice discussed in this book. We also would like to acknowledge appreciation to our students who helped us in a variety of ways to fashion the method of

presentation contained herein. We are grateful to the many users of the First Edition of *Applied Linear Statistical Models* who provided us with comments and suggestions based on their teaching with this text. We are also indebted to Professors James E. Holstein, University of Missouri, and David L. Sherry, University of West Florida, who carefully reviewed *Applied Linear Statistical Models* to provide suggestions for this volume. Dr. Robert L. Vogel assisted us diligently in the checking of the manuscript, for which we are most appreciative. Michael J. Lynn prepared the computer-generated plots using a Zeta model 3600 plotter, and George Cotsonis and Shizuki Yamamoto assisted us in the checking of calculations and in other ways. Rebecca Baggett, Marcia Pittard, and Jennifer Scott ably handled the typing and other preparation of a difficult manuscript. We are most grateful to all of these persons for their help and assistance.

Finally, our families bore patiently the pressures caused by our commitment to complete this revision. We are appreciative of their understanding.

**John Neter**  
**William Wasserman**  
**Michael H. Kutner**



---

# Contents

---

---

## 1. Some basic results in probability and statistics, 1

- 1.1 Summation and product operators, 1
- 1.2 Probability, 2
- 1.3 Random variables, 3
- 1.4 Normal probability distribution and related distributions, 6
- 1.5 Statistical estimation, 9
- 1.6 Inferences about population mean—Normal population, 10
- 1.7 Comparisons of two population means—Normal populations, 13
- 1.8 Inferences about population variance—Normal population, 16
- 1.9 Comparisons of two population variances—Normal populations, 17

---

## Part I Basic regression analysis, 21

### 2. Linear regression with one independent variable, 23

- 2.1 Relations between variables, 23
- 2.2 Regression models and their uses, 26
- 2.3 Regression model with distribution of error terms unspecified, 31
- 2.4 Estimation of regression function, 35
- 2.5 Estimation of error terms variance  $\sigma^2$ , 46
- 2.6 Normal error regression model, 48
- 2.7 Computer inputs and outputs, 51



### **3. Inferences in regression analysis, 60**

- 3.1 Inferences concerning  $\beta_1$ , 60
- 3.2 Inferences concerning  $\beta_0$ , 68
- 3.3 *Some considerations on making inferences concerning  $\beta_0$  and  $\beta_1$* , 70
- 3.4 Interval estimation of  $E(Y_h)$ , 72
- 3.5 Prediction of new observation, 76
- 3.6 Considerations in applying regression analysis, 82
- 3.7 Case when  $X$  is random, 83
- 3.8 Analysis of variance approach to regression analysis, 84
- 3.9 Descriptive measures of association between  $X$  and  $Y$  in regression model, 96
- 3.10 Computer output, 99

### **4. Aptness of model and remedial measures, 109**

- 4.1 Residuals, 109
- 4.2 Graphic analysis of residuals, 111
- 4.3 Tests involving residuals, 122
- 4.4  $F$  test for lack of fit, 123
- 4.5 Remedial measures, 132
- 4.6 Transformations, 134

### **5. Simultaneous inferences and other topics in regression analysis—I, 147**

- 5.1 Joint estimation of  $\beta_0$  and  $\beta_1$ , 147
- 5.2 Confidence band for regression line, 154
- 5.3 Simultaneous estimation of mean responses, 157
- 5.4 Simultaneous prediction intervals for new observations, 159
- 5.5 Regression through the origin, 160
- 5.6 Effect of measurement errors, 164
- 5.7 Weighted least squares, 167
- 5.8 Inverse predictions, 172
- 5.9 Choice of  $X$  levels, 175

---

## **Part II General regression and correlation analysis, 183**

### **6. Matrix approach to simple regression analysis, 185**

- 6.1 Matrices, 185
- 6.2 Matrix addition and subtraction, 190
- 6.3 Matrix multiplication, 192
- 6.4 Special types of matrices, 196

- 6.5 Linear dependence and rank of matrix, 199
- 6.6 Inverse of a matrix, 200
- 6.7 Some basic theorems for matrices, 204
- 6.8 Random vectors and matrices, 205
- 6.9 Simple linear regression model in matrix terms, 208
- 6.10 Least squares estimation of regression parameters, 210
- 6.11 Analysis of variance results, 212
- 6.12 Inferences in regression analysis, 216
- 6.13 Weighted least squares, 219
- 6.14 Residuals, 220

## **7. Multiple regression—I, 226**

- 7.1 Multiple regression models, 226
- 7.2 General linear regression model in matrix terms, 237
- 7.3 Least squares estimators, 238
- 7.4 Analysis of variance results, 239
- 7.5 Inferences about regression parameters, 242
- 7.6 Inferences about mean response, 244
- 7.7 Predictions of new observations, 246
- 7.8 An example—Multiple regression with two independent variables, 247
- 7.9 Standardized regression coefficients, 261
- 7.10 Weighted least squares, 263

## **8. Multiple regression—II, 271**

- 8.1 Multicollinearity and its effects, 271
- 8.2 Decomposition of  $SSR$  into extra sums of squares, 282
- 8.3 Coefficients of partial determination, 286
- 8.4 Testing hypotheses concerning regression coefficients in multiple regression, 289
- 8.5 Matrix formulation of general linear test, 293

## **9. Polynomial regression, 300**

- 9.1 Polynomial regression models, 300
- 9.2 Example 1—One independent variable, 305
- 9.3 Example 2—Two independent variables, 313
- 9.4 Estimating the maximum or minimum of a quadratic regression function, 317
- 9.5 Some further comments on polynomial regression, 319

## **10. Indicator variables, 328**

- 10.1 One independent qualitative variable, 328
- 10.2 Model containing interaction effects, 335

- 10.3 More complex models, 339
- 10.4 Other uses of independent indicator variables, 343
- 10.5 Some considerations in using independent indicator variables, 351
- 10.6 Dependent indicator variable, 354
- 10.7 Linear regression with dependent indicator variable, 357
- 10.8 Logistic response function, 361
- 11. Multicollinearity, influential observations, and other topics in regression analysis—II, 377**
  - 11.1 Reparameterization to improve computational accuracy, 377
  - 11.2 Problems of multicollinearity, 382
  - 11.3 Variance inflation factors and other methods of detecting multicollinearity, 390
  - 11.4 Ridge regression and other remedial measures for multicollinearity, 393
  - 11.5 Identification of outlying observations, 400
  - 11.6 Identification of influential observations and remedial measures, 407
- 12. Selection of independent variables, 417**
  - 12.1 Nature of problem, 417
  - 12.2 Example, 419
  - 12.3 All possible regression models, 421
  - 12.4 Stepwise regression, 430
  - 12.5 Selection of variables with ridge regression, 436
  - 12.6 Implementation of selection procedures, 436
- 13. Autocorrelation in time series data, 444**
  - 13.1 Problems of autocorrelation, 445
  - 13.2 First-order autoregressive error model, 448
  - 13.3 Durbin-Watson test for autocorrelation, 450
  - 13.4 Remedial measures for autocorrelation, 454
- 14. Nonlinear regression, 466**
  - 14.1 Linear, intrinsically linear, and nonlinear regression models, 466
  - 14.2 Example, 469
  - 14.3 Least squares estimation in nonlinear regression, 470
  - 14.4 Inferences about nonlinear regression parameters, 480
  - 14.5 Learning curve example, 483
- 15. Normal correlation models, 491**
  - 15.1 Distinction between regression and correlation models, 491
  - 15.2 Bivariate normal distribution, 492
  - 15.3 Conditional inferences, 496

- 15.4 Inferences on  $\rho_{12}$ , 501
- 15.5 Multivariate normal distribution, 505

### **Part III Basic analysis of variance, 515**

#### **16. Single-factor analysis of variance, 517**

- 16.1 Relation between regression and analysis of variance, 517
- 16.2 Factors, factor levels, and treatments, 520
- 16.3 Uses of analysis of variance models, 522
- 16.4 Model I—Fixed effects, 523
- 16.5 Fitting of ANOVA model, 527
- 16.6 Analysis of variance, 532
- 16.7  $F$  test for equality of factor level means, 542
- 16.8 Computer input and output for ANOVA packages, 545
- 16.9 Power of  $F$  test, 547
- 16.10 Alternative formulation of model I, 549
- 16.11 Regression approach to single-factor analysis of variance, 552

#### **17. Analysis of factor effects, 566**

- 17.1 Estimation of factor level effects, 567
- 17.2 Tukey method of multiple comparisons, 574
- 17.3 Scheffé method of multiple comparisons, 580
- 17.4 Bonferroni multiple comparison method, 582
- 17.5 Single degree of freedom tests, 584
- 17.6 Analysis of factor effects when factor quantitative, 588

#### **18. Implementation of ANOVA model, 602**

- 18.1 Planning of sample sizes with power approach, 602
- 18.2 Planning of sample sizes with estimation approach, 606
- 18.3 Planning of sample sizes to find “best” treatment, 608
- 18.4 Residual analysis, 609
- 18.5 Transformations, 615
- 18.6 Tests for equality of variances, 618
- 18.7 Effects of departures from model, 623

#### **19. Nonparametric tests, random effects model, and other topics in ANOVA—I, 636**

- 19.1 Studentized range test, 636
- 19.2 Kruskal-Wallis rank test, 638
- 19.3 Median test, 642
- 19.4 ANOVA model II—Random effects, 643

**Part IV Multifactor analysis of variance, 661****20. Two-factor analysis of variance—Equal sample sizes, 663**

- 20.1 Multifactor studies, 663
- 20.2 Meaning of model elements, 667
- 20.3 Model I (fixed effects) for two-factor studies, 682
- 20.4 Analysis of variance, 686
- 20.5  $F$  tests, 695
- 20.6 Computer input and output, 700
- 20.7 Power of  $F$  tests, 700
- 20.8 Regression approach to two-factor analysis of variance, 702

**21. Analysis of two-factor studies—Equal sample sizes, 717**

- 21.1 Strategy for analysis, 717
- 21.2 Analysis of factor effects when factors do not interact, 719
- 21.3 Analysis of factor effects when interactions important, 725
- 21.4 Analysis when one or both factors quantitative, 731
- 21.5 Implementation of two-factor ANOVA model, 735

**22. Unequal sample sizes in two-factor studies, 746**

- 22.1 Unequal sample sizes, 746
- 22.2 Use of regression approach for testing factor effects when sample sizes unequal, 747
- 22.3 Estimation of factor effects when sample sizes unequal, 754
- 22.4 Missing cells in two-factor studies, 758
- 22.5 Unequal treatment importance, 761

**23. Random and mixed effects models for two-factor studies and other topics in ANOVA—II, 773**

- 23.1 One observation per treatment, 773
- 23.2 Tukey test for additivity, 779
- 23.3 Models II (random effects) and III (mixed effects) for two-factor studies, 782
- 23.4 Analysis of variance tests for models II and III, 784
- 23.5 Estimation of factor effects for models II and III, 789

**24. Multifactor studies, 798**

- 24.1 Model I (fixed effects) for three-factor studies, 798
- 24.2 Analysis of variance, 810
- 24.3 Analysis of factor effects, 819
- 24.4 Implementation of three-factor ANOVA model, 822
- 24.5 Example of three-factor study, 823

- 24.6 Unequal sample sizes in multifactor studies, 829
- 24.7 Models II and III for three-factor studies, 831

## **25. Analysis of covariance, 844**

- 25.1 Basic ideas, 845
- 25.2 Single-factor covariance model, 848
- 25.3 Example of single-factor covariance analysis, 853
- 25.4 Single-factor covariance analysis as modification of analysis of variance, 862
- 25.5 Multifactor studies, 873
- 25.6 Additional considerations for the use of covariance analysis, 879

## **Part V Experimental designs, 891**

### **26. Completely randomized designs, 893**

- 26.1 Elements of experimental design, 894
- 26.2 Contributions of statistics to experimentation, 899
- 26.3 Completely randomized designs, 903

### **27. Randomized block designs—I, 911**

- 27.1 Basic elements, 911
- 27.2 Model, 915
- 27.3 Analysis of variance and tests, 917
- 27.4 Analysis of treatment effects, 920
- 27.5 Factorial treatments, 921
- 27.6 Planning randomized block experiments, 923
- 27.7 Analysis of aptness of model, 926
- 27.8 Regression approach to randomized block designs, 929
- 27.9 Covariance analysis for randomized block designs, 930

### **28. Randomized block designs—II, 941**

- 28.1 Binary responses for dependent variable, 941
- 28.2 Missing observations, 943
- 28.3 Random block effects, 947
- 28.4 Repeated measures designs, 948
- 28.5 Generalized randomized block designs, 952
- 28.6 Use of more than one blocking variable, 954

### **29. Nested designs—I, 963**

- 29.1 Distinction between nested and crossed factors, 963
- 29.2 Two-factor nested designs, 966
- 29.3 Analysis of variance for two-factor nested designs, 969

- 29.4 Tests for factor effects in two-factor nested designs, 975
- 29.5 Analysis of factor effects in two-factor nested designs, 977
- 29.6 Unequal nesting and replications in nested two-factor designs, 981
- 29.7 Subsampling in single-factor study with completely randomized design, 984
- 29.8 Pure subsampling in three stages, 991

### **30. Nested designs—II, 1000**

- 30.1 Rule for model development, 1000
- 30.2 Rule for finding expected mean squares, 1001
- 30.3 *Rule for finding sums of squares and degrees of freedom*, 1008
- 30.4 Crossed-nested experiment—Three factors, 1013
- 30.5 Subsampling in randomized block design, 1019
- 30.6 Two-factor experiment with repeated measures on one factor, 1021
- 30.7 Three-factor experiment with repeated measures on one factor, 1025

### **31. Latin square and related designs, 1031**

- 31.1 Basic elements, 1031
- 31.2 Model, 1039
- 31.3 Analysis of variance and tests, 1039
- 31.4 Analysis of treatment effects, 1045
- 31.5 Factorial treatments, 1046
- 31.6 Planning latin square experiments, 1047
- 31.7 Analysis of aptness of model, 1049
- 31.8 Regression approach to latin square designs, 1051
- 31.9 Missing observations, 1053
- 31.10 Additional replications with latin square designs, 1053
- 31.11 Random blocking variable effects, 1062
- 31.12 Youden and graeco-latin squares, 1064

### **Appendix tables, 1071**

### **SENIC data set, 1105**

### **SMSA data set, 1109**

### **Drug effect experiment data set, 1115**

### **Index, 1119**



## Some basic results in probability and statistics

---

This chapter contains some basic results in probability and statistics. It is intended as a reference chapter to which you may refer as you read this book. Sometimes, specific references to results in this chapter are made in the text. At other times, you may wish to refer on your own to particular results in this chapter as you feel the need.

You may prefer to scan the results on probability and statistical inference in this chapter before reading Chapter 2, or you may proceed directly to the next chapter.

### 1.1 SUMMATION AND PRODUCT OPERATORS

#### Summation operator

The summation operator  $\Sigma$  is defined as follows:

$$(1.1) \quad \sum_{i=1}^n Y_i = Y_1 + Y_2 + \cdots + Y_n$$

Some important properties of this operator are:

$$(1.2a) \quad \sum_{i=1}^n k = nk \quad \text{where } k \text{ is a constant}$$

$$(1.2b) \quad \sum_{i=1}^n (Y_i + Z_i) = \sum_{i=1}^n Y_i + \sum_{i=1}^n Z_i$$

$$(1.2c) \quad \sum_{i=1}^n (a + cY_i) = na + c \sum_{i=1}^n Y_i \quad \text{where } a \text{ and } c \text{ are constants}$$

The double summation operator  $\Sigma\Sigma$  is defined as follows:

$$(1.3) \quad \sum_{i=1}^n \sum_{j=1}^m Y_{ij} = \sum_{i=1}^n (Y_{i1} + \cdots + Y_{im}) \\ = Y_{11} + \cdots + Y_{1m} + Y_{21} + \cdots + Y_{2m} + \cdots + Y_{nm}$$

An important property of the double summation operator is:

$$(1.4) \quad \sum_{i=1}^n \sum_{j=1}^m Y_{ij} = \sum_{j=1}^m \sum_{i=1}^n Y_{ij}$$

### Product operator

The product operator  $\Pi$  is defined as follows:

$$(1.5) \quad \prod_{i=1}^n Y_i = Y_1 \cdot Y_2 \cdot Y_3 \cdots Y_n$$

## 1.2 PROBABILITY

### Addition theorem

Let  $A_i$  and  $A_j$  be two events defined on a sample space. Then:

$$(1.6) \quad P(A_i \cup A_j) = P(A_i) + P(A_j) - P(A_i \cap A_j)$$

where  $P(A_i \cup A_j)$  denotes the probability of either  $A_i$  or  $A_j$  or both occurring;  $P(A_i)$  and  $P(A_j)$  denote, respectively, the probability of  $A_i$  and the probability of  $A_j$ ; and  $P(A_i \cap A_j)$  denotes the probability of both  $A_i$  and  $A_j$  occurring.