

# Contents

<i>Detailed</i>	<i>Contents</i>	.....	xiii
<b>PART I: DATA HANDLING</b>			1
<i>Chapter 1</i>	<i>Averages and Layout</i>		3
<i>Chapter 2</i>	<i>Using Prior Knowledge</i>		28
<i>Chapter 3</i>	<i>Tables and Graphs</i>		38
<i>Chapter 4</i>	<i>Compared with What?</i>		56
<b>PART II: LAWLIKE RELATIONSHIPS</b>			65
<i>Chapter 5</i>	<i>Descriptive Relationships</i>		67
<i>Chapter 6</i>	<i>Using a Given Relationship</i>		<b>79</b>
<i>Chapter 7</i>	<i>Deriving a New Relationship</i>		95
<i>Chapter 8</i>	<i>Non-linear Relationships</i>		116
<i>Chapter 9</i>	<i>Many Variables</i>	...	135
<i>Chapter 10</i>	<i>The Emergence of Theory</i>		153
<b>PART III: STATISTICAL VARIATION</b>			167
<i>Chapter 11</i>	<i>Summary Measures</i>		<b>169</b>
<i>Chapter 12</i>	<i>Frequency Distributions</i>		187
<i>Chapter 13</i>	<i>Probability Models</i>		214
<i>Chapter 14</i>	<i>Correlation and Regression</i>		233
<i>Chapter 15</i>	<i>Multivariate Techniques</i>		<b>259</b>

<b>PART IV: SAMPLING</b>	. . . . .	283
<i>Chapter 16 Taking a Sample</i>	....	285
<i>Chapter 17 Sampling Distributions</i>	...	300
<i>Chapter 18 Statistical Inference</i>	...	314
 <b>PART V: EMPIRICAL GENERALISATION</b>	. . . . .	339
<i>Chapter 19 observation and Experimentation</i>	. . . . .	341
<i>Chapter 20 Description and Explanation</i>	. . . . .	359
 <i>List of Exercises</i>	.....	371
<i>References</i>	.....	377
<i>Name Index</i>	. . . . .	381
<i>Subject Index</i>	. . . . .	383

# Detailed Contents

<b>PART I: DATA HANDLING</b>	1
<i>Chapter 1 Averages and Layout</i>	3
1.1 Initial Visual Scanning	3
1.2 Basic Lay-out	4
1.3 Summary Figures	5
1.4 More Lay-out	7
1.5 Averages and Exceptions	9
1.6 Deviations from the Model	10
1.7 The Size of the Deviations	12
1.8 A Full Description	13
1.9 Summary	13
Exercises	14
<i>Chapter 2 Using Prior Knowledge</i>	28
2.1 New Data and Prior Knowledge	28
2.2 Irregular Deviations	29
2.3 Empirical Generalisation	31
2.4 The Beginnings of Explanation	31
2.5 Summary	32
Exercises	33
<i>Chapter 3 Tables and Graphs</i>	38
3.1 Tables to Illustrate	38
3.2 Tables for the Record	42
3.3 Graphs	43
3.4 Summary	48
Exercises	48
<i>Chapter 4 Compared with What?</i>	56
4.1 A Simplifying Concept	56
4.2 High, Low or Normal	57
4.3 The Role of Research	60
4.4 Summary	60
Exercises	61

**PART II: LAWLIKE RELATIONSHIPS**

<b>Chapter 5 Descriptive Relationships</b>	65
5.1 Linear Relationships	67
5.2 Deviations and Scatter	67
5.3 Non-linear Relationships	68
<b>5.4 The Status of Lawlike Relationships</b>	<b>70</b>
5.5 Empirical Generalisation	71
5.6 Other Things NOT Equal	72
5.7 Summary	74
Exercises	74
<b>Chapter 6 Using a Given Relationship</b>	79
6.1 Summarising the Available Data	79
6.2 Prediction and Extrapolation	79
6.3 Understanding and Theory	<b>80</b>
6.4 Technological Applications	80
6.5 Decision-making	81
6.6 The Analysis of Further Data	81
6.7 The Meaning of Failure	83
6.8 The Purpose of the Analysis	85
6.9 A Common Misuse	86
<b>6.10 Simple Prediction Again</b>	<b>89</b>
<b>6.11 The Need for Research</b>	<b>89</b>
6.~2 Summary	90
Exercises	91
<b>Chapter 7 Deriving a New Relationship</b>	95
7.1 The First-time Problem	95
7.2 A Degree of Prior Knowledge	95
7.3 The Design of the Study	96
7.4 A First Working-solution	97
7.5 The Existence of Alternative Solutions	100
7.6 A New Working-solution	<b>103</b>
7.7 Alternative Working-solutions	<b>104</b>
7.8 The Scatter of Individual Readings	<b>106</b>
7.9 Summary	<b>107</b>
Exercises	108
<b>Chapter 8 Non-linear Relationships</b>	116
8.1 Systematic Deviations	116
8.2 More Data in the Same Range	116
8.3 Choosing a Curve	119
8.4 The Fit of the Logarithmic Relationship	121
8.5 Strong Curvature	124
8.6 A Cube-root Law	126
8.7 Summary	128
Exercises	128

<b>Chapter 9 Many Variables</b> . . . . .	<b>135</b>
9.1 An Initial Analysis : <b>Apple Trees</b> . . . . .	135
9.2 A More Structured <b>Example</b> . . . . .	<b>137</b>
9.3 A Simple <b>Model</b> : Buyer Behaviour . . . . .	140
9.4 The Other Variables . . . . .	143
9.5 A <b>Breakdown in Generalisation</b> . . . . .	144
9.6 <b>Relationships in More than Two Variables</b> . . . . .	145
9.7 Correction Factors . . . . .	146
9.8 Summary . . . . .	147
<b>Exercises</b> . . . . .	147

<b>Chapter 10 The Emergence of Theory</b> . . . . .	153
10.1 <b>Different Levels of Theory</b> . . . . .	153
10.2 A Trend in Purchasing Frequencies . . . . .	155
10.3 A Theoretical Model . . . . .	156
10.4 Summary . . . . .	160
<b>Exercises</b> . . . . .	160

### **PART III: STATISTICAL VARIATION** . . . . . 167

<b>Chapter 11 Summary Measures</b> . . . . .	169
11.1 Empirical Frequency Distributions . . . . .	169
11.2 Summaries of Average Size . . . . .	171
11.3 Measures of Scatter . . . . .	174
11.4 Summary . . . . .	178
<b>Exercises</b> . . . . .	179

<b>Chapter 12 Frequency Distributions</b> . . . . .	<b>187</b>
12.1 The Normal Distribution. . . . .	<b>188</b>
12.2 The Poisson Distribution. . . . .	<b>192</b>
12.3 The Negative Binomial Distribution . . . . .	<b>194</b>
12.4 The Binomial Distribution . . . . .	196
12.5 The Beta Binomial. . . . .	199
12.6 Other Distributions . . . . .	202
12.7 Summary . . . . .	202
<b>Exercises</b> . . . . .	202

<b>Chapter 13 Probability Models</b> . . . . .	214
13.1 The Probability Concept . . . . .	214
13.2 Independent Events . . . . .	216
13.3 Stochastic Models . . . . .	218
13.4 Probability and Uncertain Events . . . . .	221
13.5 Summary . . . . .	222
<b>Exercises</b> . . . . .	222

<b>Chapter 14</b>	<b>Correlation and Regression</b>	233
14.1	<i>The Correlation Coefficient</i>	233
14.2	<i>Interpreting the Correlation Coefficient</i>	235
14.3	<i>Regression Equations</i>	238
14.4	<i>The Non-comparability of Regression Equations</i>	241
14.5	<i>Regression with One Variable Controlled</i>	245
14.6	<i>Errors in the Variables</i>	246
14.7	<i>Summary</i>	247
	<i>Exercises</i>	248
<b>Chapter 15</b>	<b>Multivariate Techniques</b>	259
15.1	<i>Multiple Regression</i>	259
15.2	<i>Component and Factor Analysis</i>	263
15.3	<i>Other Multivariate Techniques</i>	272
15.4	<i>Summary</i>	273
	<i>Exercises</i>	273
<b>PART IV: SAMPLING</b>		283
<b>Chapter 16</b>	<b>Taking a Sample</b>	285
16.1	<i>The Purpose of Sampling</i>	285
16.2	<i>Types of Statistical Sampling</i>	286
16.3	<i>Random or Probability Sampling</i>	288
16.4	<i>Technicalities of Random Sampling</i>	289
16.5	<i>The Results of Taking a Sample</i>	293
16.6	<i>Summary</i>	294
	<i>Exercises</i>	294
<b>Chapter 17</b>	<b>Sampling Distributions</b>	300
17.1	<i>Some Empirical Sampling Distributions</i>	300
17.2	<i>The Sampling Distribution of the Mean</i>	301
17.3	<i>The Difference Between Two Sample Means</i>	304
17.4	<i>Other Sampling Distributions</i>	304
17.5	<i>Non-simple Random Sampling Theory</i>	306
17.6	<i>Summary</i>	307
	<i>Exercises</i>	308
<b>Chapter 18</b>	<b>Statistical Inference</b>	314
18.1	<i>Estimation</i>	314
18.2	<i>Confidence Limits</i>	315
18.3	<i>Testing a Statistical Hypothesis</i>	318
18.4	<i>The Choice of Hypothesis</i>	320
18.5	<i>Empirical Variation</i>	322
18.6	<i>Specific Tests of Significance</i>	323
18.7	<i>Summary</i>	332
	<i>Exercises</i>	333

<b>PART V: EMPIRICAL GENERALISATION</b>	339
<b>Chapter 19 Observation and Experimentation</b>	341
19.1 Repetition . . . . .	341
19.2 Factors to Vary . . . . .	342
19.3 Statistical Surveys. . . . .	343
19.4 <b>Observational Studies</b> . . . . .	343
19.5 <b>Controlled Experimentation</b> . . . . .	344
19.6 <b>The Randomised Experiment</b> . . . . .	345
19.7 <b>The Design of Experiments</b> . . . . .	346
19.8 <b>Theoretical Norms</b> . . . . .	348
19.9 <b>Summary</b> . . . . .	349
<b>Exercises</b> . . . . .	349
 <b>Chapter 20 Description and Explanation</b>	359
2Q.1 <b>Descriptive Relationships</b> . . . . .	359
20.2 <b>Explanation</b> . . . . .	361
<b>20.3 From Facts to Theory and Back Again</b> . . . . .	362
<b>20.4 Summary</b> . . . . .	364
<b>Exercises</b> . . . . .	365
 <b>List of Exercises</b>	371
<b>References</b>	377
<b>Name Index</b>	381
<b>Subject Index</b>	383

## Preface

This is a basic textbook on how to handle and interpret statistical data. It is for anyone who comes up against statistical data in his ordinary work.

Many people feel a need to know more about statistics. They want to know how to deal with numerical data, how to judge other people's analyses, and how to choose between all the techniques that are on offer. This book is for such people: administrators and business executives, economists and technologists, social, natural, and biological scientists, and teachers and students in these areas.

There are already many books on statistics. This one differs in a number of respects. Firstly, it recognises that one is often faced with undigested tables of numbers where the author has in effect said "I do not understand these data, perhaps *you* can". The initial emphasis is therefore on how to reduce data to meaningful summaries, or how someone else should have done so in the first place. This is not adequately dealt with in most books.

Secondly, the book brings out the reasons behind statistical procedures. It emphasises *why* things are done, rather than merely giving techniques to be followed in a cookbook manner. Thirdly, although the book is primarily a basic text, it also evaluates more advanced techniques, sometimes with rather critical conclusions. These techniques are mostly employed by statistical specialists, but the general reader needs to learn something about the broad principles, so as to be able to use both the results and the specialists.

Next, many conventionally taught statistical concepts such as frequency-distributions, probability, sampling, and tests of significance are described and evaluated in the *latter* parts of the book, because they are relevant only in certain limited situations. Finally, the book recognises that results are only useful to the extent to which they *generalise*. Empirical generalisation is therefore the key concept.

The book concentrates on teaching the reader to be numerate, to see patterns and relationships that exist in numerical data and to reduce these to summaries that can readily be interpreted, used, and communicated. *The need is to let the data speak.*

Being numerate does not mean being mathematical. The use of mathematics in this book is minimal. The reader only has to be reasonably comfortable with an equation like  $y = 2x + 5$ , and occasionally with some simple algebra and a logarithm or a square root. The real role of mathematics is in the *later* stages of analysis, to help in integrating many different results. There is a natural progression from seeing the simpler pattern in one's data to developing more general models and theories, and this is touched on in various parts of the book,

To reduce the data to summary results one must see the form that the data take. The book has therefore been built around numerical examples. Some have been taken from work with which I have been personally involved and deal with ordinary, humdrum problems, not unlike those which the reader could also face.

*Data Reduction* takes a radical look at statistical methods and problems. It aims to make explicit what is best in everyday practice amongst statisticians and scientists. The basic principle throughout is that data analysis must produce results which are usable in subsequent work. Nothing could be easier in analysing new data than to see whether the new results agree with the old ones (e.g. *does*  $y = 2x + 5$  hold again?). Such use of prior results also leads to empirical generalisations; findings which are known to hold under a wide range of different conditions of observation. That is how scientific knowledge and understanding are usually generated.

### *The Structure of the Book*

The book is in five parts. Parts I and II deal with ways of seeing and communicating systematic patterns in different sets of data, primarily in terms of a single variable in Part I and relationships *between* variables in Part II.

Part III discusses the conventional statistical procedures used for describing irregular variability within a single set of readings and gives a critical evaluation of correlation and regression methods. Stochastic models are also introduced in this part.

Part IV deals with sampling and statistical inference from sample data. (Statistical teachers wishing to use the book along conventional lines may be more comfortable starting with some of the chapters in Parts III and IV.)

Part V compares the observational and experimental approaches to data collection and discusses the link between the description and explanation of observable phenomena.

Each chapter ends with exercise questions and their discussion. These fill in or elaborate various points which are either too elementary for some readers or too advanced for others and would clutter up the main text (especially at a first reading).

*Tables, Graphs and Teaching Aids*

Almost all the tables in this book have been produced on a typewriter (by Mrs. Myra Davies of Typlan Ltd.) before being reduced by one-third. They therefore take a form which can readily be produced in any report, without using a printer's different type-faces. (Two luxuries are proportional spacing, which makes the headings look more attractive but is in fact complicated to use for tables, and a simple attachment for producing bold lettering by repeated overtyping.) Mr. Robin Hipps drew the figures.

Teachers wishing copies of the original tables or figures to make transparencies or slides should write to Aske Publications at 4 Woodsyre, London SE24 6SS. The originals are about twice as large as the tables and figures in the text (50% larger linearly).

*Acknowledgments*

The preparation of this book has been greatly aided by discussions with friends and statisticians over the years, especially my colleague, Mr. G. J. Goodhardt.

For comments on numerous drafts I am indebted to many including students; and especially to Prof. M. E. Beesley, Dr. T. K. Chakrapani, Mr. P. Charlton, Dr. C. Chatfield, Mr. M. A. Collins, Mr. G. C. Naylor, Mr. S. M. Schaeffer, and Dr. H. Thomas. But I have been particularly aided by Mrs. Helen Bloom Lewis, without whom the book would have been more than twice as long and less than half as clear,

*London,*  
*January 1974*

A.S.C.E.

**Preface for the 1978 reprinting**

In this reprinting errors have been corrected. The only changes of substance from the first printing have been some rewriting of pages 212, 215 and 300.

I am indebted to readers who have pointed out mistakes, particularly Mr. J. D. F. Habbema of Erasmus University Rotterdam, Prof. W. H. Kruskal of the University of Chicago and Dr. S. Rosenbaum of the Civil Service College, London.

*London,*  
*June 1978*

A.S.C.E.