

APPLIED STATISTICS
FOR MANAGEMENT
STUDIES

THIRD EDITION

DAVID CROFT



APPLIED STATISTICS FOR MANAGEMENT STUDIES

DAVID CROFT

M.A., B.Sc., M.I.S.

Principal Lecturer in Management Statistics and Operational Research,
Management Department, Slough College of Higher Education (Constituent
of Thames Valley Regional Management Centre)

THIRD EDITION



MACDONALD AND EVANS

Macdonald & Evans Ltd
Estover, Plymouth PL6 7PZ

First published 1969
Second edition 1976
Reprinted 1979
Third edition 1983

ISBN: 0 7121 0181 9

© Macdonald & Evans 1983

This book is copyright and may not be reproduced in whole or in part (except for purposes of review) without the express permission of the publishers in writing

Printed in Great Britain by
J. W. Arrowsmith Ltd, Bristol

PREFACE TO THIRD EDITION

THIS book is intended to give an introduction to the ideas and methods of statistics to students on courses such as the Post-Graduate Diploma in Management Studies and the Work Study and Industrial Engineering Diploma. To this end I have not felt it necessary to introduce any mathematical proofs or derivations of the standard statistical formulae, and have concentrated on indicating the application rather than the theory of statistics. Experience suggests that Chapter Two in particular will be welcomed by many students, and I have tried to avoid excessive use of mathematical notation whenever possible.

I would like to express my thanks for permission to use copyright material from the Publications Board of the United Nations and the Controller of Her Majesty's Stationery Office.

I would also like to acknowledge permission to use statistical tables from the literary executor of the late Professor R. A. Fisher and Oliver & Boyd Ltd., Professor G. W. Snedecor and the Iowa State University Press and Professor A. Hald and John Wiley & Co. Inc.

Opportunity has been taken to revise and bring up to date those sections of the book dealing with current statistical practice in the United Kingdom. Some calculations have been modified to reflect the use of electronic calculators and mention has been made of the use of microcomputers in this area.

I would like to thank those colleagues and students who have made helpful comments on the earlier editions and who have encouraged me to believe that this book is meeting the need for which it was designed.

December 1982

D. C.

seemed helpful. Lastly, worked answers are now provided for many of the numerical questions set at the end of each chapter. It is hoped that these will meet the need of those students who learn best by practice.

July 1974

D. C.

CONTENTS

PREFACE TO THIRD EDITION	v
LIST OF ILLUSTRATIONS	ix
LIST OF TABLES	xi

<i>Chapter</i>		<i>page</i>
1 Introduction to statistics		1
The task of statistics · Statistics and management		
2 Introduction to statistical symbols and some mathematical reminders		6
Symbols · Mathematical notation · Calculation		
3 The Collection of Data		15
Collection · Sample Surveys · Practical Sampling · Sampling methods · Random selection methods		
4 Classification and Tabulation		32
Classification · Tabulation		
5 The presentation of statistics in diagrammatic form		39
Introduction · Pictograms and pie charts · Bar charts · Time graphs · Statistical maps		
6 Analytical graphs and diagrams		55
Logarithmic graphs · Functional graphs · Lorenz curves · Representation of frequency distributions · Planning diagrams		
7 Frequency distributions and descriptive statistical measures		69
Frequency distributions · Descriptive measures · Measures of position · Measures of spread · Measures of skewness · Population and sample statistics		
8 Weighted averages and index numbers		89
Weighted averages · Index numbers · Published index numbers · The interpretation of index numbers		
9 Probability and theoretical distributions		106
Probability · Uses of probabilities · Theoretical distribution.		

<i>Chapter</i>	<i>page</i>
10 Standard distributions and confidence intervals	125
Standard distributions · Normal curve tables	
11 Statistical significance testing	133
Testing an hypothesis · Statistical significance tests ·	
Tests on means · The " <i>t</i> "-distribution · Tests on	
proportions	
12 Statistical quality control	149
The basis of quality control · The Shewhart control	
charts · Other control charts	
13 Acceptance testing	163
The basis of sampling schemes · Designing sampling	
schemes · Multiple sampling schemes · Published	
sampling schemes	
14 The chi-squared test	176
"Goodness of fit" tests · Contingency tables	
15 The analysis of variance	188
The variance ratio test · The analysis of variance ·	
Two-factor analysis of variance · Factorial designs	
16 Correlation and linear regression	202
Correlation · Linear regression	
17 Multivariable methods	218
The multivariable problem · Factorial investigations	
· Multiple regression · Multiple correlation	
18 Time series and forecasting	228
Time series · Examination for trends · Examination	
for seasonal effects · Examination for cyclical effects ·	
Forecasting · Exponential forecasting	
APPENDIXES	
I. Selected books for further reading	243
II. Selected tables	244
III. Answers to numerical questions	251
IV. List of symbols	282
INDEX	285

LIST OF ILLUSTRATIONS

<i>Fig.</i>	<i>page</i>
1 Pictograms	40, 41
2 Pie diagrams	42
3 The indication of a broken scale	43
4 Horizontal bar chart	44
5 Multiple bar chart	44
6 Component bar chart	45
7 The choice of shadings for diagrams	46
8 Percentage component bar chart	46
9 Floating bar chart	47
10 Band chart	48
11 Linked Z-charts	49
12 Z-chart	50
13 Flow map	52
14 The use of linear and logarithmic scales	56
15 The construction of straight line graphs	57
16 The Lorenz curve	60
17 The A-B-C curve	60
18 Histogram	62
19 Frequency polygon	63
20 Cumulative frequency curve	65
21 Normal distribution	69
22 Positively skewed distribution	70
23 Bi-modal distribution	71
24 The estimation of the mode from a frequency curve	72
25 The estimation of the mode from a histogram	73
26 The estimation of the median from a cumulative frequency curve	75
27 Normal curve and standard deviation	83
28 Decision tree	114
29 The shape of binomial distributions	117
30 Operating characteristic curve	118
31 The shape of Poisson distributions	120
32 The normal distribution	122
33 The tabulated area of the normal distribution	127
34 Tabulated areas of the normal distribution	128

<i>Fig.</i>		<i>page</i>
35	The estimation of proportions using the normal distribution. Example 1	128
36	The estimation of proportions using the normal distribution. Example 2	129
37	The estimation of proportions and confidence intervals using the normal distribution. Examples 3 and 4	130
38	The use of the normal distribution in significance testing	135
39	Critical areas for two-tailed and one-tailed tests	136
40	The Shewhart control charts	153
41	The ideal operating characteristic curve	164
42	The AQL and LTPD	165
43	The derivation of the AOQL	166
44	Operating characteristic curve	167
45	Operating characteristic curves with constant sampling fraction	168
46	Operating characteristic curves with varying acceptance number	169
47	Operating characteristic curves for stages of a double sampling scheme	170
48	Sequential sampling scheme	171
49	Scatter diagram	203
50	The signs of the quadrants in the scatter diagram	203
51	Scatter diagrams for various values of r	207
52	Scatter diagrams showing perfect relationship with $r = 0$	208
53	Scatter diagrams with $r_{sp} = 1$	209
54	Linear regression line	213
55	Confidence limits for regression	215
56	Time series	228
57	The effect of a 3-point and 7-point moving average on a time series	230
58	The effect of removing seasonal effects from a time series	233
59	The simple exponential smoothing of a times series	238
60	Estimation of mode	254
61	Estimation of quartiles	256
62	Decision Tree	263

LIST OF TABLES

<i>Table</i>	<i>page</i>
I. Class boundaries for discrete data	33
II. Class boundaries for continuous data	33
III. Calculation of class intervals	34
IV. Calculation of class mid-points	35
V. Use of percentages in tables	36
VI. Construction of Z-chart	50
VII. Calculation for Lorenz curve	59
VIII. Calculation for histogram	62
IX. Calculation for cumulative frequency curve	65
X. Mode by inspection	72
XI. Estimation of the median	75
XII. Calculation of mean using coding	77
XIII. Calculation of geometric mean	78
XIV. Factors for use with mean range	80
XV. Calculation of variance and standard deviation	84
XVI. Calculation of price relatives	90
XVII. Price relatives—change of base	91
XVIII. Calculation of index number from relatives	92
XIX. Calculation of index number using the weighted geometric mean of relatives	94
XX. Calculation of aggregative index	94
XXI. Retail price index weightings	99
XXII. Pay-off table for horse bet	111
XXIII. Probability of various demands	112
XXIV. Pay-off tables for 3 and 4 units stocked	112, 113
XXV. Expectation of profit for various strategies	113
XXVI. Calculation for operating characteristic curve	118
XXVII. Use of the Poisson distribution	121
XXVIII. Use of the Poisson distribution	121
XXIX. Constants for standard distributions	123
XXX. Significance levels	137
XXXI. Calculation of population variance estimate	140
XXXII. Paired comparison test	144
XXXIII. Factors for use with \bar{w} on control charts	156
XXXIV. Control limits for defects	160

<i>Table</i>	<i>page</i>
XXXV. Calculation for operating characteristic curve	167
XXXVI. Expected frequencies from Poisson distribution	177
XXXVII. Calculation of χ^2	177
XXXVIII. Expected frequencies by use of normal distribution	179
XXXIX. Calculation of χ^3	179
XL. Contingency table	180
XLI. Expected frequencies in contingency table	181
XLII. Calculation of χ^2	182
XLIII. Observed and expected frequencies 2×2 table	184
XLIV. Calculation of χ^2 for 2×2 table with Yates' correction	184
XLV. Anovar table	191
XLVI. Calculation for one-way Anovar	192
XLVII. Anovar table—one-way	193
XLVIII. Data for two-way Anovar	195
XLIX. Calculation for two-way Anovar	195
L. Anovar table—two-way	196
LI. Anovar table with replication	197
LII. Latin square	199
LIII. Calculation of correlation coefficient	206
LIV. Calculation of rank correlation coefficient	210
LV. Anovar for regression equation	213
LVI. Test of significance of extra factors	221
LVII. Calculation of moving averages	231
LVIII. Calculation of seasonal effects	232
LIX. Exponential smoothing of time series	237

CHAPTER ONE

INTRODUCTION TO STATISTICS

The task of statistics

Development of statistics as an analytical tool

THE techniques of preparing, analysing and presenting data which are now included under the title of "Statistics" have been developed in many different fields of study. This has had the advantage of bringing to the subject a variety of approach which has prevented any over-formalisation, but has unfortunately led to a multiplicity of different terms and symbols having the same meaning. Stemming from this varied background is the characteristic of statistics that the methods pioneered in one field of application are often developed and standardised in another.

Some of the main contributions to statistics may be indicated by a short history of the subject. The word "statistics" appears to have been first used in the late eighteenth century, to refer to the verbal description of the resources and political organisation of European states then being published. The word gradually transferred to numerical data of this type, and in the form of "official" or "government" statistics is still very important. At the same time, information was being collected on the behaviour of prices, and measures such as "index numbers" were developed. These were the subject of considerable theoretical discussion in the late nineteenth century, and again in the 1920s and are still capable of arousing controversy.

Another line of statistical development started with the discussions on gambling in the seventeenth century which led to the basic ideas of probability. With the growth of measurement as a tool of scientific inquiry, the results of probability problems were applied to the study of errors in measurements and so to the formulation of the theoretical distributions of measurements. In the exact sciences, where careful control of experimental procedure was leading to a high degree of reproducibility of results, these statistical results were little used, but by the late nineteenth century the study of agriculture and biology by experiment and

measurement was producing a mass of variable results, and these could only be dealt with by statistical methods.

Much that is still of great value in statistics was developed at this stage and the association of statistics with agricultural science led to the establishment by the 1920s of many of the major methods of statistical analysis. It was also becoming realised that the same methods were suitable, with slight modifications, for the analysis of any measurement showing a mixture of controlled and uncontrolled variability, and statistical results began to be used in sociology, psychology and economic studies, and in industrial and commercial applications. Typical of this was the development of sampling for opinion polls and market research just prior to the Second World War.

The late 1920s saw the development of a purely industrial statistical technique, that of "statistical quality control," but the major advances in this field came under the impetus of the mass-production of armaments during the Second World War. Since 1950 it has become more and more obvious that industrial and commercial data are well suited to analysis by statistical methods, and new methods and applications are continually being reported. Certain fields such as market research are by now largely based on the use of statistical methods. Similarly in a recent survey of operational research scientists in which they were asked which analytical techniques they made use of regularly, only statistical analysis was named by more than 50%.

The latest development has been the increasing availability of electronic computers and calculating machines to handle computational work. This has led to a renewed interest in methods which were known theoretically, but were not thought feasible because of the amount of computation they required. Some of these are particularly interesting in the study of industrial and commercial problems involving many uncontrolled variables. Even the simpler techniques become much easier to handle, and the growth of the hand-held electronic calculator industry is providing tools which allow "on the spot" analysis and will give a further boost to the use of statistical methods as a routine procedure in business.

Modern uses of statistics

Statistics in the modern world have an application in any situation in which numerical information is handled, and especially

where the data is of a variable nature. Statistical results and procedures are used at all stages of the processing of data. When the data is to be collected, statistics help in identifying the most efficient methods to use, whether experiment or survey. When the data is to be analysed, statistical measures summarise the information collected, and statistical tests examine it for significance. When the results are reported, statistical statements present the conclusions unambiguously and statistical diagrams display them in visual form.

It is important to realise that the use of good statistical methodology can lead to a considerable increase in the value of numerical information, and that its absence is often an indication that the information may be poor or even misleading.

Statistics and management

Statistics and scientific management

The art of management is the art of taking decisions, and the basic necessity for correct decision-taking is good information. It is becoming increasingly recognised that the more information there is available the better the decision taken. What is termed "scientific management" is basically management on the basis of precise numerical information, and to try to use such numerical information without using statistics to aid in its preparation is to fail to use one of the major tools of efficient management. In fact even when not recognised in its own right statistics is an essential part of most modern management services. In market research, in work study, in production control and planning, in quality control, in stock control, in sales analysis and forecasting, in project planning and in operational research there is a core of statistical methodology. Increasingly, also, the nature of decision-taking and the inherent element of probability is becoming recognised, and statistics is providing a vocabulary in which risks and confidence can be recognised and measured.

The statistician in management

Given this picture of management as a profession increasingly concerned with numerical information and analysis, the statistician has an important place as an adviser. The fear is often expressed

that the "experts" in measurement and numerical manipulation will supplant the manager by presenting results which remove doubt and the need for decision-taking. This is far from true, and as more information becomes available in any situation, then the critical decisions become more clearly focused and often more difficult to make. Where information is incomplete or inaccurate an incorrect decision can often escape notice, and an inadequate manager fails to realise his inadequacies. The role of the manager is likely to grow more important and more demanding as he is able to draw on more and more expert information.

In this situation the statistician appears as a provider of information and advice, but not as a taker of decisions. He should be consulted at the early stages of the collection of data to ensure that the most efficient methods are used. He will, of course, be mainly concerned with the analysis of the information collected, and he will present the results in such a way that their content and significance is clear. He may indicate the implications of the results, but taking action to avoid or realise these implications is the responsibility of the manager.

Statistics in management education

As the emphasis on scientific management increases, so does the demand for management education, and the interested manager is told to study a wide and increasing range of subjects. Of these, statistics is usually regarded as a basic necessity and there are two main reasons why this is so.

First, the manager will be taking decisions and assessing probabilities, whether consciously or not. Statistical method can often provide some degree of measurability in an apparently unmeasurable situation. The manager will sometimes also have to reduce numerical data to a pattern of information suiting his particular needs, and again statistical methods are appropriate. The need to use statistical method is likely to grow with the expansion of the facilities available for computation on large computers via remote terminals, and on programmable desk calculators. These facilities will include libraries of standard calculations, and their very ease of use places more importance on the correct choice of data to feed into them and the correct interpretation of the results obtained.

Secondly, the manager will wish to know when the statistician

can make a useful contribution to the analysis of a situation, and what limitations there are on the use of statistical methods. He may also need to assess the validity of conclusions presented to him and to argue with their originator.

From this it follows that there is a dual requirement in the statistical education of a manager. He should be able to use those methods which are appropriate to non-expert use, and he should appreciate the power and basis of statistics and should know when to call in a more expert analyst. There is no better way to achieve the first of these requirements than the actual practical use of statistical methods, and some arithmetical labour is well rewarded by a better knowledge of the methods involved.

CHAPTER TWO

INTRODUCTION TO STATISTICAL SYMBOLS AND SOME MATHEMATICAL REMINDERS

Symbols

Use of symbols

THE use of symbols instead of words in statistical formulae produces a considerable saving in space and time, and there seems little point in trying to avoid the use of such convenient aids to understanding. The symbols used in this book have been chosen in an attempt to avoid confusion, and their meaning will be described when they are introduced in the text. The symbols used correspond as far as possible to those in general use in the United Kingdom and the U.S.A. Unfortunately there is no universal set of symbols in statistics, and the reader is advised to check if similar symbols have similar meanings when referring to other statistical literature.

Special conventions

Two particular conventions will be adhered to throughout this book.

(a) Original quantities will be symbolised by capital letters, *e.g.* X , while derived quantities will be symbolised by the corresponding small letter, *e.g.* x . If the distinction is not necessary, then the small letters only are used.

(b) Statistics derived from samples will be symbolised by English letters, *e.g.* m , s , while the corresponding population statistics will be symbolised by the equivalent Greek letter, *e.g.* μ , σ . The use of these conventions will be further explained as they are required.

Suffixes

Suffixes will be used to distinguish between quantities of a similar type, such as repeated measurements of the same quantity: