MASTERING

STATISTICS

WITH YOUR MICROCOMPUTER

MACMILLAN MASTER SERIES

Banking
Basic English Law
Basic Management
Biology
British Politics
Business Communication
Chemistry
COBOL Programming
Commerce
Computer Programming
Computers
Data Processing
Economics
Electrical Engineering
Electronics
English Grammar
English Language
English Literature
French
French 2
German

Hairdressing
Italian
Keyboarding
Marketing
Mathematics
Modern British History
Modern World History
Nutrition
Office Practice
Pascal Programming
Physics
Principles of Accounts
Social Welfare
Sociology
Spanish
Spanish 2
Statistics
Statistics with Your Microcomputer
Study Skills
Typewriting Skills
Word Processing

MASTERING STATISTICS WITH YOUR MICROCOMPUTER

CONALL BOYLE

M
MACMILLAN

Text © Conall Boyle 1986
Figures © Macmillan Education Ltd 1986
Cartoons © Jim Bray 1986

All rights reserved. No reproduction, copy or transmission of this publication may be made without written permission.

No paragraph of this publication may be reproduced, copied or transmitted save with written permission or in accordance with the provisions of the Copyright Act 1956 (as amended).

Any person who does any unauthorised act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

First published 1986

Published by
MACMILLAN EDUCATION LTD
Houndmills, Basingstoke, Hampshire RG21 2XS
and London
Companies and representatives
throughout the world

Typeset by TecSet Ltd, Sutton, Surrey

British Library Cataloguing in Publication Data
Boyle, Conall
Mastering statistics with your microcomputer. —
(Macmillan master series)

1. Statistics—Data processing
2. Microprocessors

I. Title
519.5'028'5404 QA276.4

ISBN 978-0-333-39172-3 ISBN 978-1-349-18211-4 (eBook)
DOI 10.1007/978-1-349-18211-4

*To my boys Alex and Marcus
in the hope that their joy in finding out will never be diminished*

CONTENTS

<i>Acknowledgements</i>	xi
<i>Introduction</i>	xiii

PART I COLLECTING THE DATA

1 Data base	1.1 Data: What is it? Where can you find it?	3
	1.2 Setting up the data base: house prices	4
	1.3 Looking ahead	7
2 Store data on your computer	2.1 Use of the input command	9
	2.2 Use of READ . . . DATA	10
	2.3 Read data into an array (subscripted variable)	12
	2.4 Enrichment activity - two extra tasks	13
3 Manual recording of data	3.1 Tally mark chart	15
	3.2 The histogram	19
	3.3 The stem and leaf plot	20
	3.4 Further developments of the stem and leaf plot	21
	3.5 Single digit or category plot: dotbox	22
4 Types of data and scales of measurement		24

PART II SORT AND DISPLAY THE DATA

5 Sorting data into order		29
6 Find the hinges	6.1 Counts up and down	33
	6.2 Hinges: the market posts along the data	34
	6.3 5-number summary	36

CONTENTS

	6.4	Discrete and classificatory data – summaries	37
	6.5	What do hinges tell us about the data?	42
	6.6	Appendix: program to calculate hinge points in discrete data	42
7 Display the data: box-plots	7.1	Box and whisker plots	44
	7.2	Hints for plotters	46
	7.3	Further variations on the box-plot	47
	7.4	Polygons and cumulative frequency polygons	49

PART III SUM UP AND EXPLAIN THE DATA

8 Middle values	8.1	Central measures	55
	8.2	Isn't central measure just the average?	55
	8.3	Central measures: names	55
	8.4	Calculation methods	56
	8.5	Resistancy	58
	8.6	What are these central measures measuring?	58
	8.7	Applications of central measures	61
9 Spread values	9.1	A measure of spread	62
	9.2	Measure of spread: names	62
	9.3	Measure of spread: calculations	62
	9.4	Measure of spread: experiments	65
	9.5	Resistancy of the measures?	66
	9.6	Size of the measure?	66
	9.7	Size of standard denation for varying N	67
	9.8	Why ($-\bar{x}$) for mean deviation and standard deviation?	67
	9.9	Measures of spread: applications	69

10 Skewness and re-expression	10.1	Shape of the distribution on the stem and leaf plot	71
	10.2	Skewness: names	71
	10.3	Identifying skewness	71
	10.4	Skewness and central values	72
	10.5	Dealing with skewness: re-expression	73
	10.6	Staircase of re-expressions	75
	10.7	Application of re-expression	77

PART IV COMPARING DATA

11 Comparing distributions	11.1	Comparing new with old	82
	11.2	Multiple box-plots	84
	11.3	Notched box-plot	85
	11.4	Formula for notched box-plot	86
12 Index numbers	12.1	Definition	88
	12.2	Population	89
	12.3	Simple index	91
	12.4	Aggregate index	91
	12.5	Weighted index	92
	12.6	Weighted price-relative index	93
	12.7	Index numbers: should they be scrapped?	94
	12.8	Applications	94

PART V RELATIONSHIPS IN DATA

13 Tables	13.1	Taming the clutter: coded tables	101
	13.2	Polishing the table	103
14 Correlations	14.1	Relationships	108
	14.2	Plotting relationships	108
	14.3	Scatterplots on the computer	110
	14.4	Fitting a formula	112
	14.5	Testing the fit: residuals	114
	14.6	Improving the fit: better values for a and b	117
	14.7	Improving the fit: method of least squares	118

CONTENTS

	14.8	Method of least squares by formula	119
	14.9	Correlation coefficient	120
15 Time series	15.1	What is a time series?	122
	15.2	Data	123
	15.3	Plotting the data	124
	15.4	Smoothing the data	125
	15.5	Double smoothing	127
	15.6	Multiple smoothing	129
	15.7	Residuals: the 'rough'	129

PART VI FURTHER THOUGHTS

16 Mistakes, errors and surveys	16.1	How to prevent or correct mistakes in your surveys	135
	16.2	Error control	138
17 Where from? Where to	17.1	How far have you come?	140
	17.2	The next step: more statistics	141
	17.3	Statistics on the computer	144
	17.4	The longer term	146
18 What is different about this book?	18.1	What's wrong with current teaching of statistics?	147
	18.2	What should a course in statistics contain?	148
	18.3	Self-confidence	150

<i>Index</i>			151
--------------	--	--	-----

ACKNOWLEDGEMENTS

I would like to thank many people for helping me to achieve this book.

John Bibby of the Open University for encouraging me to try EDA.

My colleagues at Birmingham Polytechnic, especially Bob Farmer for his encouragement, and Roger Ball for his tolerance; Andrew King and Gordon Kelly for their help.

The students on quantity surveying and estate management degree courses who so willingly tried out earlier drafts of this book, and from whom I learned so much. We lecturers must never forget that without our students we are nothing.

My wife Nesta who understands that doing this book is important to me.

I would also like to thank the Consumers' Association, publishers of *Which?*, for permission to use the diagram in Figure 7.4; and the editor of the *Guardian* for permission to reprint an extract in Chapter 12.

CONALL BOYLE

INTRODUCTION

This is a learning-by-doing book. With the aid of your microcomputer, it aims to introduce you to the delights of statistics. You will also learn a great deal about using your computer to produce results. In each chapter in Parts I-V there are a number of *Tasks* for you to do. Completing these tasks will enable you to master statistics.

INTRODUCTION

THE EQUIPMENT YOU WILL NEED FOR THESE LEARNING TASKS

- A microcomputer. Any sort will do so long as it can run BASIC programs.
- A cassette recorder attached to your micro for program and data recording (or a disk unit if you are fortunate).
- A supply of blank cassettes (or disks).
- A printer is *very* useful, but you can get by without one.

A method of recording your results – a ring binder or a notebook – is also needed. As you complete each task, you should write it up and file it away. This will give you a record of the progress you are making, and will be handy to refer to later.

THE PREVIOUS KNOWLEDGE YOU REQUIRE FOR THIS BOOK

- Mathematics – simple arithmetic is all you really need. This is a statistics course based on understanding, not algebra.
- Computer programming – since you have a computer, you should have learned a little BASIC by now. At least a smattering of BASIC is needed before you start on this course. Your instruction manual will fill in any gaps in your knowledge of programming.

ORGANISING FOR LEARNING

As this book is designed for self-learning, I think you will find it most helpful to get yourself organised. If you draw up a check-list of tasks, you can record when you start and finish each one. You can also set yourself target dates by which you hope to finish each chapter and part.

An *assignment* in this book could consist of a completed set of tasks from a single part. So your first assignment could be to complete the tasks in Part I.

I do hope you will enjoy yourself exploring and mastering statistics with the help of your microcomputer!