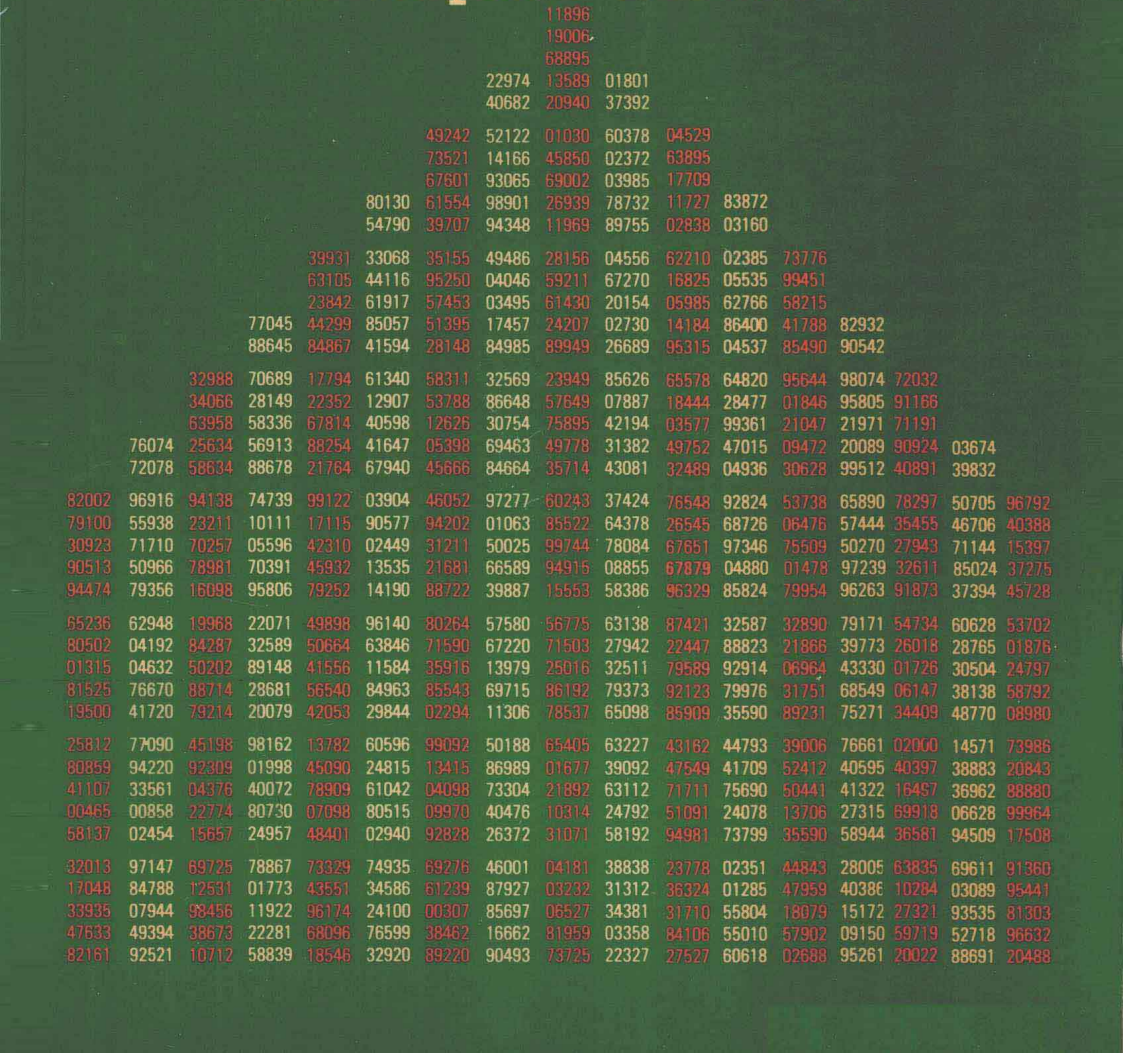
Statistics for technology

A COURSE IN APPLIED STATISTICS

Third edition

Christopher Chatfield



Statistics for technology

Statistics for technology

A Course in Applied Statistics

Christopher Chatfield

Senior Lecturer in Statistics,
Bath University, UK

THIRD EDITION

LONDON NEW YORK
CHAPMAN AND HALL

First published 1970
by Penguin Books Ltd
1975 reprint published by
Chapman and Hall Ltd
11 New Fetter Lane, London EC4P 4EE
Second edition 1978
Third edition 1983
Published in the USA by
Chapman and Hall
733 Third Avenue, New York NY10017

© 1983 Christopher Chatfield

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

ISBN 0 412 25340 2

This paperback edition is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, re-sold, hired out, or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

All rights reserved. No part of this book may be reprinted, or reproduced or utilized in any form or by any electronic, mechanical or other means, now known or hereafter invented, including photocopying and recording, or in any information storage and retrieval system, without permission in writing from the Publisher.

British Library Cataloguing in Publication Data

Chatfield, Christopher
Statistics for technology.—3rd ed.—(Science paperbacks; 114)
1. Mathematical statistics
I. Title
519.5'0246 QA276
ISBN 0-412-25340-2

Library of Congress Cataloging in Publication Data

Chatfield, Christopher
Statistics for technology.
(Science paperbacks; 114)
Bibliography: p.
Includes index.
1. Engineering—Statistical methods. I. Title.
II. Series.
TA340.C45 1983 519.5'02462 83-2108
ISBN 0-412-25340-2 (pbk.)

试读结束：需要全本请在线购买： www.ertongbook.com

Contents

Preface 11

Part One Introduction

1 **Outline of statistics** 15

2 **Simple ways of summarizing data** 20
2.1 Introduction 20
2.2 Graphical methods 21
2.3 Summary statistics 28

Part Two Theory

3 **The concept of probability** 37
3.1 Probability and statistics 37
3.2 Some definitions 38
3.3 Types of events 40
3.4 Permutations and combinations 49
3.5 Random variables 51

4 **Discrete distributions** 56
4.1 The discrete probability distribution 56
4.2 The binomial distribution 57
4.3 The binomial model 59
4.4 Types of distribution 64
4.5 The mean and variance of the binomial distribution 64
4.6 The Poisson distribution 69
4.7 Bivariate discrete distributions 75

5 **Continuous distributions** 81
5.1 Definitions 81
5.2 The mean and variance of continuous distributions 86
5.3 The normal distribution 87
5.4 Uses of the normal distribution 92
5.5 Normal probability paper 95
5.6 The exponential distribution 98
5.7 Bivariate continuous distributions 103

6 Estimation 106
6.1 Point and interval estimates 106
6.2 Properties of the expected value 107
6.3 The sampling distribution of $\bar{x}$ 111
6.4 The sampling distribution of s^2 116
6.5 Some properties of estimators 118
6.6 General methods of point estimation 121
6.7 Interval estimation 126

7 Significance tests 134
7.1 Introduction 134
7.2 Tests on a sample mean 140
7.3 Comparing two sample means 143
7.4 The *t*-test applied to paired comparisons 147
7.5 The χ^2 goodness-of-fit test 148
7.6 The *F*-test 155
7.7 Distribution-free or non-parametric tests 157
7.8 Power and other considerations 158

8 Regression and correlation 166
8.1 Scatter diagram 166
8.2 Curve fitting 167
8.3 Regression 171
8.4 Confidence intervals and significance tests in linear regression 174
8.5 The coefficient of determination 177
8.6 Multiple and curvilinear regression 178
8.7 Orthogonal polynomials 180
8.8 The design of regression experiments 185
8.9 The correlation coefficient 185
8.10 Estimating the regression lines 191
8.11 The bivariate normal distribution 194
8.12 Interpretation of the correlation coefficient 196

Part Three Applications

9 Planning the experiment 203
9.1 Preliminary remarks 203
9.2 Measurements 204
9.3 The propagation of error 206
9.4 Improving precision with series and parallel arrangements 215
9.5 Combining dissimilar estimates by the method of least squares 216

10 The design and analysis of experiments – 1 Comparative experiments 224
10.1 Some basic considerations in experimental design 224
10.2 A mathematical model for simple comparative experiments 226

10.3 The number of replications 227
10.4 Randomization 230
10.5 The analysis of a randomized comparative experiment 231
10.6 The range test 235
10.7 One-way analysis of variance 237
10.8 Follow-up study of the treatment means 241
10.9 Verifying the model 243
10.10 The randomized block experiment 244
10.11 Two-way analysis of variance 248
10.12 Latin squares 252
10.13 Balanced incomplete block designs 253

11 The design and analysis of experiments – 2 Factorial experiments 257
11.1 Introduction 257
11.2 The advantages of complete factorial experiments 258
11.3 The design of complete factorial experiments 260
11.4 The analysis of a complete factorial experiment 263
11.5 Follow-up procedure 269
11.6 The 2^n factorial design 271
11.7 Fixed effects and random effects 276
11.8 Other topics 277
11.9 The examination of residuals 279
11.10 Determination of optimum conditions 280
11.11 Summary 285

12 Quality control 288
12.1 Acceptance sampling 288
12.2 Operating characteristic curve 289
12.3 Types of sampling schemes 293
12.4 Rectifying schemes 296
12.5 The military standard plan 297
12.6 Sampling by variables 298
12.7 Practical problems 298
12.8 Process control 299
12.9 Control charts for samples 301
12.10 Cusum charts 306
12.11 Prediction, system identification and control 312

13 Life testing 319
13.1 Problems in measuring reliability 319
13.2 The mathematical distribution of failure times 321
13.3 The exponential distribution 323
13.4 Estimating the conditional failure rate 324
13.5 The Weibull distribution 327

Appendix A The relationships between the normal, χ^2, t- and F-distributions 332

Appendix B Statistical tables 335
Table 1. Areas under the normal curve 335
Table 2. Percentage points of Student's t-distribution 336
Table 3. Percentage points of the χ^2 distribution 337
Table 4. Upper percentage points of the F-distribution 338
Table 5. Values of e^{-x} 341
Table 6. Percentage points of the distribution of the Studentized range 342
Table 7. Random numbers 344

Appendix C Further reading 346

Appendix D Some other topics 351
D.1 Calculating the mean and standard deviation of a frequency distribution 351
D.2 Interpretation of the sample standard deviation 354
D.3 How to round numbers 355
D.4 Stem-and-leaf plots and box plots 360
D.5 Estimating and testing a proportion 363

Appendix E Some general comments on tackling statistical problems 365
E.1 Preliminary questions 365
E.2 Collecting the data 366
E.3 Preliminary data analysis 367
E.4 Choice of a more elaborate method 369
E.5 Computer programs 370
E.6 Using a library 371
E.7 Presenting the results 371

Answers to exercises 373

Index 377

'It's perfectly intelligible,' the captain said, in an offended tone, 'to anyone that understands such things.'

Preface

This book provides an introduction to statistics, with particular emphasis on applications in the applied sciences and engineering. The book may be used as a text for a basic course in statistics or for self-tuition. Although the book was originally intended for 'service' courses to scientists and engineers, I have been pleased to find that the book has also been widely used for applied statistics courses for students in mathematics and statistics departments. Although 'technology' was an 'in' word when the book was written, it may be that a better title today would be 'A Course in Applied Statistics'.

The book is divided into three sections. Part One includes an introduction and some material on descriptive statistics. Part Two deals with the theory of probability and statistics, while Part Three considers some applications, including the design and analysis of experiments, quality control and life-testing. The reader who is specifically interested in one of these last two topics can proceed directly from Chapter 7 to Chapter 12 or 13 as appropriate.

The favourable reaction to the first and second editions has prompted me to make relatively few changes for the third edition. The references have been updated throughout, and Sections 2.1 and 6.3 have been rewritten. In the second edition Sections 12.11 and 13.4 were rewritten. I have also clarified and updated the text with many minor alterations, but without changing the page numbering up to page 346. This should be helpful to teachers who are used to earlier editions. It should also keep costs down and avoid introducing new typographical errors.

Appendix C, giving further references, has been completely rewritten again. A new Appendix D was introduced in the second edition describing some miscellaneous topics which were omitted from the first edition. In this edition I have added a new Section D.4 on stem-and-leaf plots and box plots. I have also added a completely new Appendix E in which I make some general comments on how to tackle statistical problems. I strongly recommend that the reader reads this Appendix

试读结束：需要全本请在线购买： www.ertongbook.com

after taking a conventional introductory course. It includes remarks on topics such as processing the data, preliminary data analysis, computer programs, using a library, and writing a report. These vital topics are often omitted from conventional courses.

I have kept the mathematics throughout the book as simple as possible; an elementary knowledge of calculus plus the ability to manipulate algebraic formulae are all that is required. I have tried to introduce the theory of statistics in a comprehensible way without getting too involved in mathematical details. A few results are stated without proof where this is unlikely to affect the student's comprehension. However, I have tried to explain carefully the basic concepts of the subject, such as probability and sampling distributions; these the student must understand. The worst abuses of statistics arise in the 'cook-book' approach to the subject where scientists try to analyse their data by substituting measurements into statistical formulae which they do not understand.

Many readers will have access to a computer, or at least to a microcomputer or sophisticated pocket calculator. The statistical software which is readily available has taken much of the drudgery out of statistics. There are some remarks on computer programs in Appendix E.5. I have amended the text in a number of places to make it clear that the reader need no longer worry too much about some computational details, although there is much to be said for working through some of the methods in detail at least once in order to understand fully what is going on.

I am grateful to many people for constructive comments on earlier editions and I am always pleased to hear from any reader with new ideas.

I am indebted to Biometrika trustees for permission to publish extracts from *Biometrika Tables for Statisticians* (Appendix B, Table 6) and to Messrs Oliver and Boyd Ltd, Edinburgh for permission to publish extracts from *Statistical Tables for Biological, Agricultural and Medical Research* (Appendix B, Table 7).

The quotations are, of course, from the works of Charles Lutwidge Dodgson, better known as Lewis Carroll.

CHRISTOPHER CHATFIELD
Bath University
December, 1982

Part one
Introduction

'Surely,' said the governor, 'Her Radiancy would admit that ten is nearer to ten than nine is – and also nearer than eleven is.'

Chapter 1
Outline of statistics

Statistics is an increasingly important subject which is useful in many types of scientific investigation. It has become the science of collecting, analysing and interpreting data in the best possible way. Statistics is particularly useful in situations where there is *experimental uncertainty* and may be defined as 'the science of making decisions in the face of uncertainty'. We begin with some scientific examples in which experimental uncertainty is present.

Example 1

The thrust of a rocket engine was measured at ten-minute intervals while being run at the same operating conditions. The following thirty observations were recorded (in newtons $\times 10^5$):

999·1	1003·2	1002·1	999·2	989·7	1006·7
1012·3	996·4	1000·2	995·3	1008·7	993·4
998·1	997·9	1003·1	1002·6	1001·8	996·5
992·8	1006·5	1004·5	1000·3	1014·5	998·6
989·4	1002·9	999·3	994·7	1007·6	1000·9

The observations vary between 989·4 and 1014·5 with an average value of about 1000. There is no apparent reason for this variation which is of course small compared with the absolute magnitude of each observation; nor do the variations appear to be systematic in any way. Any variation in which there is no pattern or regularity is called *random variation.* In this case if the running conditions are kept uniform we can predict that the next observation will also be about a thousand together with a small random quantity which may be positive or negative.

Example 2

The numbers of cosmic particles striking an apparatus in forty consecutive periods of one minute were recorded as follows.

3	0	0	1	0	2	1	0	1	1
0	3	4	1	2	0	2	0	3	1
1	0	1	2	0	2	1	0	1	2
3	1	0	0	2	1	0	3	1	2

The observations vary between zero and four, with zero and one observed more frequently than two, three and four. Again there is experimental uncertainty since we cannot exactly predict what the next observation would be. However, we expect that it will also be between zero and four and that it is more likely to be a zero or a one than anything else. In Chapter 4 we will see that there is indeed a pattern in this data even though individual observations cannot be predicted.

Example 3

Twenty refrigerator motors were run to destruction under advanced stress conditions and the times to failure (in hours) were recorded as follows.

104·3	158·7	193·7	201·3	206·2
227·8	249·1	307·8	311·5	329·6
358·5	364·3	370·4	380·5	394·6
426·2	434·1	552·6	594·0	691·5

We cannot predict exactly how long an individual motor will last, but, if possible, we would like to predict the pattern of behaviour of a batch of motors. For example we might want to know the over-all proportion of motors which last longer than one week (168 hours). This problem will be discussed in Chapter 13.

When the scientist or engineer finishes his education and enters industry for the first time, he must be prepared to be faced frequently with situations which involve experimental uncertainty. The purpose of this book is to provide the scientist with methods for treating these uncertainties. These methods have proved to be very useful in both industry and research.

试读结束：需要全本请在线购买： www.ertongbook.com