THE CAMBRIDGE DICTIONARY OF STATISTICS

THE CAMBRIDGE DICTIONARY OF Statistics

Second Edition

B. S. EVERITT
Institute of Psychiatry, King's College, University of London

CAMBRIDGE UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 1998
Reprinted 1999
Second edition 2002

Printed in the United Kingdom at the University Press, Cambridge

Typeface 9/12pt Times. *System* 3B2 [KW]

A catalogue record for this book is available from the British Library

ISBN 0 521 81099 x hardback

To the memory of my dear sister Iris

Preface

The Cambridge Dictionary of Statistics aims to provide students of statistics, working statisticians and researchers in many disciplines who are users of statistics with relatively concise definitions of statistical terms. All areas of statistics are covered, theoretical, applied, medical, etc., although, as in any dictionary, the choice of which terms to include and which to exclude, is likely to reflect some aspects of the compiler's main areas of interest, and I have no illusions that this dictionary is any different. My hope is that the dictionary will provide a useful source of reference for both specialists and non-specialists alike. Many definitions necessarily contain some mathematical formulae and/or nomeclature, others contain none. But the difference in mathematical content and level among the definitions will, with luck, largely reflect the type of reader likely to turn to a particular definition. The non-specialist looking up, for example, **Student's t-tests** will hopefully find the simple formulae and associated written material more than adequate to satisfy their curiosity, while the specialist seeking a quick reminder about **spline functions** will find the more extensive technical material just what they need.

The dictionary contains approximately 3000 headwords and short biographies of more than 100 important statisticians (fellow statisticians who regard themselves as 'important' but who are *not* included here should note the single common characteristics of those who are). Several forms of cross-referencing are used. Terms in *slanted roman* in an entry appear as a separate headword, although headwords defining relatively commonly occurring terms such as **random variable**, **probability**, **distribution**, **population**, **sample**, etc., are *not* referred to in this way. Some entries simply refer readers to another entry. This may indicate that the terms are synonyms or, alternatively, that the term is more conveniently discussed under another entry. In the latter case the term is printed in *italics* in the main entry.

Entries are in alphabetical order using the letter-by-letter rather than the word-by-word convention. In terms containing numbers or Greek letters, the numbers or corresponding English word are spelt out and alphabetized accordingly. So, for example, 2×2 table is found under **two-by-two table**, and α-trimmed mean, under **alpha-trimmed mean**. Only headings corresponding to names are inverted, so the entry for William Gossett is found under **Gosset, William**. but there is an entry under **Box–Müller transformation** *not* under **Transformation, Box–Müller**.

For those readers seeking more detailed information about a topic, many entries contain either a reference to one or other of the texts listed later, or a more specific reference to a relevant book or journal article. (Entries for software contain the appropriate address.) Additional material is also available in many cases in either the *Encyclopedia of Statistical Sciences*, edited by Kotz and Johnson, or the *Encyclopedia of Biostatistics*, edited by Armitage and Colton, both published by Wiley. Extended biographies of many of the people included in this dictionary can also be found in these two encyclopedias and also in *Leading Personalities in Statistical Sciences* by Johnson and Kotz published in 1997 again by Wiley.

Lastly and paraphrasing Oscar Wilde 'writing one dictionary is suspect, writing two borders on the pathological'. But before readers jump to an obvious conclusion I would like to make it very clear that an anorak has never featured in my wardrobe.

B. S. Everitt, 1998

Preface to second edition

In this second edition of the *Cambridge Dictionary of Statistics* I have added approximately 500 new entries and taken the opportunity to correct and clarify a number of the previous entries. I have also added biographies of important statisticians whom I overlooked in the first edition and, sadly, I have had to include a number of new biographies of statisticians who have died since the publication of the first edition in 1998.

I would like to thank Professor Pak Sham for providing a number of the entries on genetics, Professor David Hand for putting me straight about data mining and related techniques, Professor David Finney for pointing out problems with a number of the definitions in the first edition and Dr Phillip Sedgwick for leading me to the correct (I hope) definition of censored observations.

B. S. Everitt, 2001

Acknowledgements

Firstly I would like to thank the many authors who have, unwittingly, provided the basis of a large number of the definitions included in this dictionary through their books and papers. Next thanks are due to many members of the 'allstat' mailing list who helped with references to particular terms. I am also extremely grateful to my colleagues, Dr Sophia Rabe-Hesketh and Dr Sabine Landau, for their careful reading of the text and their numerous helpful suggestions. Lastly I have to thank my secretary, Mrs Harriet Meteyard, for maintaining and typing the many files that contained the material for the dictionary and for her constant reassurance that nothing was lost!

Sources

Altman, D.G. (1991) *Practical Statistics for Medical Research*, Chapman and Hall, London. (SMR)

Chatfield, C. (1996) *The Analysis of Time Series: An Introduction*, 5th edition, Chapman and Hall, London. (TMS)

Evans, M., Hastings, N. and Peacock, B. (2000) *Statistical Distributions*, 3rd edition, Wiley, New York. (STD)

Krzanowski, W.J. and Marriot, F.H.C. (1994) *Multivariate Analysis, Part 1*, Edward Arnold, London. (MV1)

Krzanowski, W.J. and Marriot, F.H.C. (1995) *Multivariate Analysis, Part 2*, Edward Arnold, London. (MV2)

McCullagh, P.M. and Nelder, J.A. (1989) *Generalized Linear Models*, 2nd edition, Chapman and Hall, London. (GLM)

Rawlings, J.O. (1988) *Applied Regression Analysis*, Wadsworth Books, California. (ARA)

Stuart, A. and Ord, K. (1994) *Kendall's Advanced Theory of Statistics, Volume 1*, 6th edition, Edward Arnold, London. (KA1)

Stuart, A. and Ord, K. (1991) *Kendall's Advanced Theory of Statistics, Volume 2*, 5th edition, Edward Arnold, London. (KA2)