Applied Multivariate Data Analysis

Applied Multivariate Data Analysis

Second Edition

Brian S. Everitt
Institute of Psychiatry, King's College London, UK

and

Graham Dunn
School of Epidemiology and Health Sciences, University of Manchester, UK

John Wiley & Sons, Ltd

First published in Great Britain in 2001 by Arnold
This impression printed by Hodder Education,
a part of Hachette Livre UK,
338 Euston Road, London NW1 3BH

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN 978-0-4707-1117-0

8 9 10

Typeset in 10/12pt Times by Academic & Technical Typesetting. Bristol

Contents

Preface

The majority of data sets collected by researchers in all disciplines are multivariate. In a few cases it may be sensible to isolate each variable and study it separately, but in most instances all the variables need to be examined simultaneously in order to fully grasp the structure and key features of the data. For this purpose, one or other method of multivariate analysis may be helpful, and it is with such methods that this book is largely concerned.

Multivariate analysis includes methods both for describing and exploring data and for more formal inferential procedures. The aim of all the techniques is, in a general sense, to display or extract the signal in the data in the presence of noise, and to find out what the data show us in the midst of their apparent chaos.

We have made many changes from the first edition of our book, including a separate chapter on correspondence analysis, a section on neural networks for classification, and discussion of extensions of the generalized linear model to situations involving multiple response variables – for example, repeated measures studies. In addition, the graphical techniques chapter has been completely rewritten and many new graphical methods described. Finally, more – and, we hope – better, examples illustrating techniques are to be found in all chapters. As with the first edition, we have aimed the book both at students on statistics courses and at applied researchers dealing with multivariate data. Readers need to have some background in statistics, perhaps of the kind delivered by an introductory course covering estimation, inference, regression, analysis of variance and so on. The main mathematical requirement is a degree of familiarity with matrix algebra, although much of the more technical material is confined to tables so that the less mathematical reader will often be able to follow the discussion to some extent.

In the first part of the book (Chapters 2 to 6) we concentrate largely on what might loosely be described as the exploratory multivariate techniques; often these are primarily graphical in nature, and the graphical display of multivariate