DATA ANALYSIS IN SPORT

PETER O'DONOGHUE
AND LUCY HOLMES

ROUTLEDGE STUDIES IN
SPORTS PERFORMANCE ANALYSIS

ROUTLEDGE

Making sense of sports performance data can be a challenging task but is nevertheless an essential part of performance analysis investigations. Focusing on techniques used in the analysis of sports performance, this book introduces the fundamental principles of data analysis, explores the most important tools used in data analysis, and offers guidance on the presentation of results.

The book covers key topics such as:

- the purpose of data analysis, from statistical analysis to algorithmic processing;
- commercial packages for performance and data analysis, including Focus, Sportscode, Dartfish, Prozone, Excel, SPSS and Matlab;
- effective use of statistical procedures in sports performance analysis;
- analysing data from manual notation systems, player tracking systems and computerised match analysis systems;
- creating visually appealing 'dashboard' interfaces for presenting data;
- assessing reliability.

The book includes worked examples from real sport, offering clear guidance to the reader and bringing the subject to life. This book is invaluable reading for any student, researcher or analyst working in sports performance or undertaking a sport-related research project or methods course.

Peter O'Donoghue is Reader and Discipline Director for Performance Analysis at the Cardiff School of Sport, Cardiff Metropolitan University, UK. He is a member of the International Society of Performance Analysis of Sport, editor of the *International Journal of Performance Analysis of Sport*, and editor of the book series *Routledge Studies in Sports Performance Analysis*.

Lucy Holmes is Lecturer in Performance Analysis and Programme Director for the MSc Performance Analysis at the Cardiff School of Sport, Cardiff Metropolitan University, UK. She is currently Lead Performance Analyst for Hockey Wales.

Sport Science / Research Methods

Cover image: © Getty

ISBN 978-0-415-73984-9

9 780415 739849

Routledge
Taylor & Francis Group

www.routledge.com

an informa business

Routledge titles are available as eBook editions in a range of digital formats

DATA ANALYSIS IN SPORT

PETER O'DC
AND LUCY

dge

DATA ANALYSIS IN SPORT

PETER O'DONOGHUE AND LUCY HOLMES

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2015
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
711 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

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
O'Donoghue, Peter.
Data analysis in sport / Peter O'Donoghue, Lucy Holmes.
pages cm. — (Routledge studies in sports performance analysis)
Includes bibliographical references and index.
1. Sports—Physiological aspects. 2. Sports sciences—Statistical methods. 3. Sports sciences—Research—Methodology. 4. Performance—Statistical methods. 5. Performance—Research—Methodology. I. Holmes, Lucy (Teacher in performance analysis) II. Title.
RC1235.O357 2015
613.7072'7—dc23
2014017975

ISBN: 978-0-415-73983-2 (hbk)
ISBN: 978-0-415-73984-9 (pbk)
ISBN: 978-1-315-81635-7 (ebk)

Typeset in Melior and Univers
by FiSH Books Ltd, Enfield

DATA ANALYSIS IN SPORT

Making sense of sports performance data can be a challenging task but is nevertheless an essential part of performance analysis investigations. Focusing on techniques used in the analysis of sport performance, this book introduces the fundamental principles of data analysis, explores the most important tools used in data analysis, and offers guidance on the presentation of results.

The book covers key topics such as:

- the purpose of data analysis, from statistical analysis to algorithmic processing;
- commercial packages for performance and data analysis, including Focus, SportsCode, Dartfish, Prozone, Excel, SPSS and Matlab;
- effective use of statistical procedures in sport performance analysis;
- analysing data from manual notation systems, player tracking systems and computerised match analysis systems;
- creating visually appealing 'dashboard' interfaces for presenting data;
- assessing reliability.

The book includes worked examples from real sport, offering clear guidance to the reader and bringing the subject to life. This book is invaluable reading for any student, researcher or analyst working in sport performance or undertaking a sport-related research project or methods course.

Peter O'Donoghue is Reader and Discipline Director for Performance Analysis at the Cardiff School of Sport, Cardiff Metropolitan University, UK. He is a member of the International Society of Performance Analysis of Sport, editor of the *International Journal of Performance Analysis of Sport* and editor of the book series *Routledge Studies in Sport Performance Analysis*.

Lucy Holmes is Lecturer in Performance Analysis and Programme Director for the MSc Performance Analysis at the Cardiff School of Sport, Cardiff Metropolitan University, UK. She is currently Lead Performance Analyst for Hockey Wales.

Routledge Studies in Sports Performance Analysis

Series Editor

Peter O'Donoghue

Cardiff Metropolitan University

Routledge Studies in Sports Performance Analysis is designed to support students, lecturers and practitioners in all areas of this important and rapidly developing discipline. Books in the series are written by leading international experts in sports performance analysis and cover topics including match analysis, analysis of individual sports and team sports, technique analysis, data analytics, performance analysis for high performance management, and various methodological areas. Drawing on the very latest research, and introducing key concepts and best practice, the series meets a need for accessible, up-to-date texts at all levels of study and work in performance analysis.

Available in this series:

An Introduction to Performance Analysis of Sport
Peter O'Donoghue

Data Analysis in Sport
Peter O'Donoghue and Lucy Holmes

To our parents

PREFACE

This second book in the *Routledge Studies in Sports Performance Analysis* series deals with data analysis in sports performance analysis. There are many existing textbooks on quantitative analysis, qualitative analysis and mixed methods in sport as well as more generally. The rationale for this book was that sports performance data present some unique challenges for data analysis that cannot expect to be covered by more general data analysis books. The book covers reasonably advanced material but does not go into the more advanced areas of data analytics that will be covered in a separate book within the series. The book will be of interest to performance analysts with the ambition to undertake data analysis tasks that go beyond the standard data analyses outputs provided by commercial match analysis packages. Such analysts could be working with elite athletes or in educational environments. There is a specialist modelling, profiling and statistics module within the MSc Sport Performance Analysis at the authors' university that is supported by the book. The book also serves as a reference text and, therefore, the chapters cover different aspects of data analysis rather than forming a strict sequence of chapters that must be read in order from beginning to end. The intended readership is expected to be performance analysts who are already capable of developing and using performance analysis systems. A pre-requisite to using the current textbook is that the analysts are competent users of commercial match analysis packages and are capable of developing and using sports specific systems within these packages. They should be capable of using the standard outputs of commercial packages including event lists, frequency tables of events, interactive video review facilities and highlight movie creation. There is

an alternative type of reader who does not engage in system development or operation but who may be tasked with performing sophisticated analyses of sports performance data gathered by others.

Existing textbooks in sports performance analysis cover some of the standard analysis tasks performed in the area. For example, the *Routledge Handbook of Sports Performance Analysis* (McGarry *et al.*, 2013) covers profiling and reviews work done using artificial neural networks. *Research Methods in Sports Performance Analysis* (O'Donoghue, 2010) is aimed at students undertaking research projects within the final year of their undergraduate degree or masters programme and covers qualitative methods, non-parametric statistical tests and analysis of reliability that are relevant to sports performance research. *Notational Analysis of Sport* (Hughes and Franks, 2004a) contains chapters on reliability, profiling and modelling. *The Essentials of Performance Analysis of Sport* (Hughes and Franks, 2008) contains chapters on probabilistic modelling in sport and reliability. However, there are data analysis techniques that are done by performance analysts that are not covered in any textbooks. Furthermore, there are data pre-processing tasks that are necessary before many sports performance data can be analysed for reliability, and other pre-processing tasks performed on some data prior to applying statistical analyses. The current book fills some of these gaps but the authors would not claim that the book covers all of the advanced data analysis techniques that could be applied to sports performance data. For example, there is a chapter on Matlab. Matlab is a data analysis and modelling package about which several volumes of user guides have been written. Similarly, an analyst with the ability to use 10 per cent of the functionality of Microsoft Excel will be very valuable to the athletes they work with.

The book contains nine chapters, the first two of which are a natural start to the book. There was a choice of ordering of the remaining chapters, which cover different types of analysis. The first chapter discusses the nature of data and information and some general principles of data analysis. The second chapter discusses the standard data analysis features of commercial video analysis packages that readers are already expected to be familiar with. The reason for including this chapter is that many readers may use one particular package without being aware of facilities or different ways of abstracting behaviour that are used within alternative packages.

Microsoft Excel is a widely used data analysis package that many students and analysts have installed on the computers that they use. The general purpose nature of Microsoft Excel means that it can be used for more types of analyses than the analysis functions provided by commercial match analysis packages. The commercial match analysis packages provide the most fundamental analysis features for tagged video analysis. However, they cannot be expected to include all possible analysis functions for all possible sports it could be used for. This requires analysts to export data to other packages for more specialist analysis. The reason why a general purpose data analysis package like Microsoft Excel can be used is that it is such a popular package: Microsoft has been able to include wide-ranging features for statistical analysis, text processing, mathematical functions and graphics. Event lists can be exported from commercial match analysis packages and summarised using pivot tables and other functions of Microsoft Excel.

Once you have collected your data, you need to be able to present this information in a usable format for the intended viewer. Dashboards have been used in business intelligence for many years; the transfer of this knowledge to sport with the growth of the data collected was inevitable. Being able to communicate quickly and efficiently to your players, coaches and other personnel the information you have collected is vital. This presentation then becomes the first access point that allows further investigation using the performance analysis packages.

The higher-end versions of the Sportscode package provide a Statistical Window which currently has no equivalent in the alternative packages. The Statistical Window is a grid of cells that contain scripts that are programmed to calculate and display information. The scripts access data from the timeline and the information computed can be sent to the code window or an output window. There are many Sportscode users who do not use Statistical Windows within their systems. Script programming requires similar skills to other types of computer programming. A minority of skilled analysts are capable of developing systems that include Statistical Windows. The fifth chapter of the book is aimed at aspiring analysts who wish to develop this ability. It is an introduction that takes readers through three examples ranging from simple percentage calculation to a more complex example of automatically updating the score of a tennis match. Those interested in script programming should support the reading of this chapter by developing scripts of use to their own systems. Those with an ability to programme scripts can potentially work

as consultants developing systems for use by other analysts working with athletes and squads.

Chapter 6 discusses the analysis of player tracking data in team sports. A variety of player tracking systems is used, ranging from relatively low-cost GPS systems to high-cost image processing systems that are used in professional soccer clubs. The player tracking systems do have analysis facilities that provide useful outputs. However, there are tactical behaviours of teams that could potentially be recognised from player tracking data. Chapter 6 proposes a five stage process of developing such algorithms using challenging in soccer as an example. Algorithms that recognise tactical aspects of collective behaviour in team sports are potentially beneficial to squads. The efficiency of an automated process allows for feedback about such behaviours to be made using quantitative information or relevant video sequences.

Matlab is a programming language and environment used for data analysis, modelling and simulation. It allows complex data structures to be created and analysed. Matlab has great potential for the analysis of sports performance data. Chapter 7 uses an example of analysing player tracking data to automatically identify where defences apply pressure, backup and cover according to Olsen's (1981) principles. A second example is a simulation system for the 2014 FIFA World Cup. There are toolkits for artificial neural networks, image processing and other advanced analyses that the book does not go into. The chapter gives readers an introduction that will allow them to start developing programming skills in Matlab. Scientific research into sports performance particularly can benefit from the use of Matlab.

Chapter 8 covers statistical analysis of sports performance data. This is a topic covered in other textbooks. However, the pre-processing of sports performance data prior to statistical analysis is of particular interest to the chapter. Many students are more than capable of selecting the correct statistical procedure to use and use statistical packages to produce statistical results. However, they sometimes have difficulty transforming data into a form that can be analysed using statistics packages. Chapter 8 discusses the use of pivot tables and other types of pre-processing that transform data into a form where they can be entered into SPSS for statistical analysis.

The final chapter covers reliability assessment in sports performance analysis. This is another area that has been covered in depth elsewhere.

However, as with statistical analysis, data need to be transformed into a form that allows reliability statistics to be calculated. The chapter shows how timelines and event lists from independent observations can be combined to allow reliability statistics to be determined. The reliability of both input data and output information can be evaluated. An example of this is reliability assessment for commercial match analysis packages in soccer. Another issue covered in the chapter is the reliability of match statistics provided on official tournament websites.

ACKNOWLEDGEMENTS

We'd like to thank the commercial organisations who granted permission for us to use screenshots of different parts of their systems. The SPSS company granted permission to include SPSS outputs in the current book. Microsoft granted permission to use screenshots of popup windows of their system that have been used. Prozone granted permission for us to reanalyse some data used in a 2006 reliability study so that additional reliability statistics could be reported. Dartfish granted permission for us to show screenshots of different parts of the Dartfish package.

We'd also like to thank William Bailey, Josh Wells and Simon Whitmore from Routledge for their assistance during the planning and developing of this book.

Thank you all.

CONTENTS