



STATISTICS

FOR SPORT AND EXERCISE STUDIES

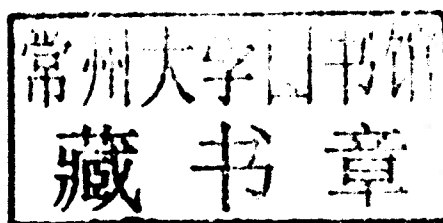
AN INTRODUCTION

PETER O'DONOGHUE

STATISTICS FOR SPORT AND EXERCISE STUDIES

An introduction

PETER O'DONOGHUE



First published 2012
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
711 Third Avenue, New York, NY 10017

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

© 2012 Peter O'Donoghue

The right of Peter O'Donoghue to be identified as author of this work has been asserted by him in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

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 or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

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

Statistics for sport and exercise studies : an introduction /
edited by Peter O'Donoghue.

p. cm.

1. Sports sciences. 2. Sports—Statistics. 3. Exercise—Statistics.

I. O'Donoghue, Peter.

GV558.S73 2012

613.7'1—dc23

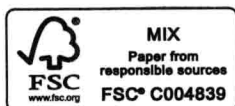
2011040739

ISBN: 978-0-415-59556-8 (hbk)

ISBN: 978-0-415-59557-5 (pbk)

ISBN: 978-0-203-13350-7 (ebk)

Typeset in Zapf Humanist 601 BT
by Swales & Willis Ltd, Exeter, Devon



Printed and bound in Great Britain by
TJ International Ltd, Padstow, Cornwall

STATISTICS FOR SPORT AND EXERCISE STUDIES

Statistics for Sport and Exercise Studies guides the student through the full research process, from selecting the most appropriate statistical procedure, to analysing data, to the presentation of results, illustrating every key step in the process with clear examples, case studies and data taken from sport and exercise settings.

Every chapter includes a range of features designed to help the student grasp the underlying concepts and relate each statistical procedure to their own research project, including definitions of key terms, practical exercises, worked examples, clear summaries and guides to further reading and resources. The book also offers an in-depth and practical guide to using SPSS in sport and exercise research, the most commonly used data analysis software in sport and exercise departments. In addition, a companion website includes downloadable data sets and work sheets for use in or out of the classroom.

This book is a complete, user-friendly and easy-to-read introduction to the use of statistical tests, techniques and procedures in sport, exercise and related subjects.

Visit the companion website for this book at www.routledge.com/cw/odonoghue/

Peter O'Donoghue is Reader and Discipline Director for Performance Analysis in the Cardiff School of Sport, Cardiff Metropolitan University, UK. Peter started his career as a computer scientist before moving into sport and exercise science in 1997. Since then he has been active in performance analysis of sport and is Chair of the International Society of Performance Analysis of Sport.

To Mum

PREFACE

Statistical analysis has been and continues to be an area where many students have difficulties. This textbook, like the many other statistics textbooks, can only form part of the solution to this problem. The author has read other statistics textbooks, particularly those that are specifically aimed at sport and exercise sciences. While there is a rationale for the current textbook, the author wishes to recognize the outstanding work of Ntoumanis (2001), Hinton (2004), Hinton *et al.* (2004), Fallowfield *et al.* (2005) and Newell *et al.* (2010). These are excellent textbooks with different emphases on statistical procedures, the use of statistical packages and statistics within a research context. The key for successful development in the area of statistics is engagement with the area. This textbook is designed to assist students who are prepared to do the work necessary to develop the knowledge, intellectual and practical skills of the area.

In designing this textbook, the author considered how statistics are used within research and identified three things that need to be done: (a) select the most appropriate statistical procedures for the given research design; (b) apply the procedures to research data; and (c) report the results of data analysis. Therefore, each of the chapters covering statistical procedures follows a structure based on these phases aided by the use of a sport or exercise science example. The book uses SPSS (SPSS: An IBM Company, Armonk, NY) as an example computerized statistical analysis package. SPSS data sheets are provided on an accompanying website to allow students to try the examples for themselves and compare their results with those in the book. The exercises with accompanying data and solutions are also found on the website.

While engaging in these exercises is strongly encouraged, the author also had to consider where and when readers might be reading the textbook. It is not always convenient for readers to be logged onto a computer that has the SPSS package installed. Some may prefer to read through chapters before engaging in the exercises. Therefore, the book was written so that it could be used without the accompanying resources. However, the full potential of the book will only be realized by those who do engage in the practical exercises.

Developing the book has also been a learning experience for the author who has enjoyed the discussions he has had with colleagues about the examples that can be used and types of analyses that the book should promote. The book has also made a couple of contributions that may be original. The reason that the author says 'may be original' is because the two developments seem so obvious that somebody somewhere must have made such proposals

before. One is an alternative to the Bonferroni adjustment used to restrict the experiment-wise α level. The author has always been uncomfortable with the Bonferroni adjustment leading to p values of 1.000 being reported because probabilities of greater than 1 have been calculated! The alternative approach proposed in Chapter 11 ensures an α level for individual pairwise comparisons that leads precisely to the intended experiment-wise α level. The second contribution is post hoc testing for the chi square test of independence. Where an association between two categorical variables measured at more than two levels is found, the author proposes testing each nominal value against the others put together to determine precisely where significance occurs.

This book is a reference book and is not intended to be read from cover to cover in sequential order. Indeed, the placement of some chapters was a difficult decision to make. In particular, there is a case for covering reliability before the main descriptive and inferential procedures. The first chapter is a general introduction to the area of statistics within the context of quantitative research. The second chapter is more of a housekeeping chapter providing guidance on how to use the book. Those who feel they cannot deal with equations and formulae should read this chapter and hopefully find that they are not so difficult. Chapter 3 covers the all-important descriptive statistics before Chapter 4 considers the use of standardized scores to assist interpretation of statistics. Chapters 5 and 6 provide essential background material on probability and probability distributions respectively which are pre-requisite to Chapter 7 where hypothesis testing is introduced. Chapters 8 and 9 cover correlation and regression. Chapters 11 to 15 cover inferential statistical procedures before chapters 16 to 18 cover some of multivariate procedures that few sport and exercise science students would cover before masters level. Chapter 19 was an interesting one for the author to write because it highlights that different disciplines use the word 'reliability' to mean different things. Chapter 20 is a very brief introduction to statistical power analysis and is heavily influenced by the work of Murphy *et al.* (2009). Those particularly interested in statistical power are referred to the more in-depth coverage provided by Murphy *et al.*'s (2009) book.

ILLUSTRATION CREDITS

The figures listed below are reprinted courtesy of International Business Machines Corporation, © SPSS, Inc., an IBM Company. SPSS was acquired by IBM in October, 2009.

- 2.1 The variable definition form in SPSS
- 2.2 Datasheet viewer window in SPSS
- 2.3 The split file facility in SPSS
- 2.4 Computing a new variable in SPSS
- 2.5 The SPSS output viewer
- 2.6 Menu navigation in SPSS
- 2.7 Bivariate correlations pop-up window
- 3.1 Variable view
- 3.2 Frequencies pop-up window
- 3.3 Cross-tabulation pop-up window
- 3.4 Cell options for cross-tabulation
- 3.6 Statistics for ordinal variables
- 3.7 Descriptive pop-up window
- 3.8 Options within descriptive statistics
- 3.9 Exploring variables in SPSS
- 3.10 Requesting percentiles when exploring variables
- 3.13 Comparing means for groups identified by an independent variable
- 3.14 Options for comparing means
- 3.15 Sorting cases into descending order of population
- 3.16 Explore with a factor
 - 6.5 The Explore facility in SPSS
 - 6.6 Selecting normality plots and tests when exploring variables
 - 6.9 Descriptive statistics pop-up window
 - 6.10 Selecting analysis of distribution in terms of skewness and kurtosis
- 6.12 The Transform-Compute facility in SPSS
 - 8.2 The Chart Builder pop-up window
 - 8.3 The SPSS Chart Editor
 - 8.4 Pop-up window for editing properties of chart elements
 - 8.5 Bivariate correlations pop-up window
 - 8.6 Pop-up window for partial correlations
 - 9.3 Chart Editor

- 9.4 Chart Properties
- 9.5 Scatter plot with regression line
- 9.6 Linear regression pop-up window
- 9.7 Statistics pop-up window for linear regression
- 9.8 The Save pop-up window for linear regression
- 9.9 Linear regression pop-up window using more than one independent variable
- 9.10 Statistics pop-up for linear regression
- 9.11 More precise inspection of values in SPSS output
- 9.12 Options pop-up for linear regression
- 10.1 One-sample t-test pop-up window in SPSS
- 10.2 Options pop-up window for the one-sample t-test
- 10.3 Pop-up window for the independent samples t-test
- 10.5 Pop-up window for the paired samples t-test
- 11.1 Pop-up window for the one-way ANOVA
- 11.2 Options for the one-way ANOVA test
- 11.3 Post hoc test selection
- 11.4 Defining a within subjects factor when performing a repeated measures ANOVA test
- 11.5 Identifying the repeated measurements of a dependent variable
- 11.6 Options pop-up for the repeated measures ANOVA
- 11.9 Setting up an ANCOVA test in SPSS
- 11.10 Saving predicted values and residuals when performing an ANCOVA test
- 12.1 SPSS window for univariate ANOVA tests
- 12.2 Saving predicted and residual values when performing a univariate ANOVA test
- 12.3 Applying an ANOVA test to the residuals to test homogeneity of variances
- 12.4 Options pop-up window for requesting homogeneity tests when performing univariate ANOVA tests
- 12.5 Options pop-up window for univariate ANOVA tests
- 12.7 Defining a within-within design
- 12.8 Repeated measures of our conceptual dependent variable are assigned to factor level combinations
- 12.9 Saving predicted and residual values when performing an ANOVA test including two within-subjects effects
- 12.10 Selecting options within a repeated measures ANOVA test
- 12.12 Identifying a within-subjects factor within a mixed ANOVA
- 12.13 Defining a mixed ANOVA test
- 12.14 Selecting options in a mixed ANOVA test
- 13.1 Multivariate pop-up window in SPSS
- 13.2 Options pop-up window within the Multivariate facility of SPSS
- 13.3 Post hoc test selection for MANOVA tests
- 13.4 Use of the repeated measures ANOVA facility in SPSS to perform a repeated measures MANOVA
- 13.5 Linking variables (columns) in the SPSS datasheet to repeated measures of conceptual dependent variables

- 14.1 The pop-up window to compare two independent samples using a non-parametric test
- 14.3 Pop-up window for two related samples tests to be done non-parametrically
- 14.5 Pop-up window for the Kruskal–Wallis H test
- 14.7 Pop-up window for the Friedman test in SPSS
- 15.1 Chi Square pop-up window
- 15.3 Crosstabulation pop-up window
- 15.4 Statistics pop-up window for cross-tabulations
- 16.1 The discriminant function analysis pop-up window
- 16.3 The pop-up window for binary logistic regression
- 17.1 Hierarchical cluster analysis pop-up window
- 18.1 The Factor Analysis pop-up window
- 19.6 Reliability Analysis pop-up window in SPSS

ACKNOWLEDGEMENTS

I'd like to express my sincere thanks to all the people who helped me during the writing and production of this book. In particular, I'd like to thank Richard Neil, the Discipline Director of Undergraduate Research Methods at University of Wales Institute Cardiff (Cardiff Metropolitan University), who developed some of the examples used in this textbook and proofread the chapters that use them. I'm very grateful to Laurence Llewelyn, a former colleague, from whom I learned some new tricks in SPSS and statistical analysis in general. This textbook has been motivated by the excellent and exciting work that has been done in statistics in recent years at Cardiff Metropolitan University and I'd like to express my thanks to the members of the team I have been part of: Owen Thomas, Steve-Mark Cooper, Ian Bezodis, Michael Hughes, Jon Oliver, Paul Smith, Gemma Robinson, Genevieve Williams, Kylie Wilson, Ray Ponting, Deborah Welford, Ian Mitchell, Rhodri Lloyd, Joseph Esformes, Marianne Gittoes, Cassie Wilson, Nic James, Karianne Backx, Christian Edwards, David Wasley and Kieran Kingston. I'd also like to thank Josh Wells and Simon Whitmore from Routledge for their assistance during the planning and developing of this book.

Thank you all.

xxiv

acknowledgements

CONTENTS

<i>List of figures</i>	ix
<i>List of tables</i>	xiii
<i>Preface</i>	xix
<i>Illustration credits</i>	xxi
<i>Acknowledgements</i>	xxiv
1 Data, information and statistics	1
2 Using this book	19
3 Descriptive statistics	33
4 Standardized scores	57
5 Probability	71
6 Data distributions	88
7 Hypothesis testing	111
8 Correlation	135
9 Linear regression	149
10 t-Tests	171
11 Analysis of variances	191
12 Factorial ANOVA	218
13 Multivariate ANOVA	242
14 Non-parametric tests	267
15 Chi square	285
16 Statistical classification	298
17 Cluster analysis	313

18	Data reduction using principal components analysis	326
19	Reliability	337
20	Statistical power	368
	<i>References</i>	379
	<i>Index</i>	386

FIGURES

2.1	The variable definition form in SPSS	24
2.2	Datasheet viewer window in SPSS	25
2.3	The split file facility in SPSS	26
2.4	Computing a new variable in SPSS	26
2.5	The SPSS output viewer	27
2.6	Menu navigation in SPSS	28
2.7	Bivariate correlations pop-up window	29
3.1	Variable view	36
3.2	Frequencies pop-up window	36
3.3	Cross-tabulation pop-up window	37
3.4	Cell options for cross-tabulation	38
3.5	Using a map to present descriptive statistics	39
3.6	Statistics for ordinal variables	41
3.7	Descriptive pop-up window	45
3.8	Options within descriptive statistics	45
3.9	Exploring variables in SPSS	47
3.10	Requesting percentiles when exploring variables	47
3.11	Box and whiskers plot for population	49
3.12	Box and whiskers plot for capacity of the largest stadium in the city	49
3.13	Comparing means for groups identified by an independent variable	50
3.14	Options for comparing means	50
3.15	Sorting cases into descending order of population	52
3.16	Explore with a factor	52
3.17	Clustered box and whisker plots for population when type of city (major) is used as a factor	53
3.18	Capacity of largest stadium	54
4.1	Box and whiskers plot for the percentage of centre passes leading to goal in netball matches between different qualities of team	65
5.1	Tree diagram to represent a multistep experiment	74
5.2	Venn diagram showing experimental outcomes within each event	80
5.3	A probabilistic model for winning a game in tennis	84
6.1	Uniform continuous probability distribution	92
6.2	Raw scores, z-scores and T-scores for a normally distributed variable	94
6.3	Distribution of men's finishing times in the 2011 London marathon	96

6.4	Non-normal distributions	97
6.5	The Explore facility in SPSS	98
6.6	Selecting normality plots and tests when exploring variables	99
6.7	Normal Q-Q plot produced by SPSS	99
6.8	Box and whiskers plot for the percentage of points where the first serve was in	100
6.9	Descriptive statistics pop-up window	103
6.10	Selecting analysis of distribution in terms of skewness and kurtosis	103
6.11	Box and whiskers plot showing positive skewness	105
6.12	The Transform-Compute facility in SPSS	106
6.13	t distributions	106
6.14	F distributions	107
6.15	Chi square distributions	108
7.1	Spreadsheet design for a population of 10,000	118
7.2	95 per cent confidence intervals determined from different samples	121
7.3	Estimated sampling distribution for the mean of the control and experimental group	124
7.4	Comparing distributions of sample means	126
7.5	Tests of single variables	128
7.6	Tests of relationships between pairs of variables	128
7.7	Tests of differences between independent groups	129
7.8	Tests of differences between related groups	130
7.9	Predictive modelling techniques	131
8.1	Different types of correlation	135
8.2	The Chart Builder pop-up window	138
8.3	The SPSS Chart Editor	139
8.4	Pop-up window for editing properties of chart elements	140
8.5	Bivariate correlations pop-up window	141
8.6	Pop-up window for partial correlations	144
9.1	Regression line $y = a + b.x$	150
9.2	Relation between 800m PB and 1,500m PB	153
9.3	Chart Editor	154
9.4	Chart Properties	155
9.5	Scatter plot with regression line	156
9.6	Linear Regression pop-up window	157
9.7	Statistics pop-up window for linear regression	158
9.8	The Save pop-up window for linear regression	159
9.9	Linear Regression pop-up window using more than one independent variable	163
9.10	Statistics pop-up for linear regression	164
9.11	More precise inspection of values in SPSS output	166
9.12	Options pop-up for linear regression	167
10.1	One-sample t-test pop-up window in SPSS	174
10.2	Options pop-up window for the one-sample t-test	174
10.3	Pop-up window for the independent samples t-test	179
10.4	Presentation of descriptive results relating to the independent t-test	182
10.5	Pop-up window for the paired samples t-test	186
10.6	Sit ups performed in 60s under using different methods of self-talk	188



11.1	Pop-up window for the one-way ANOVA	194
11.2	Options for the one-way ANOVA test	194
11.3	Post hoc test selection	195
11.4	Defining a within-subjects factor when performing a repeated measures ANOVA test	204
11.5	Identifying the repeated measurements of a dependent variable	205
11.6	Options pop-up for the repeated measures ANOVA	206
11.7	Percentage of netball match spent performing high-intensity activity	209
11.8	Relationship between reaction time and 60m sprint time	210
11.9	Setting up an ANCOVA test in SPSS	212
11.10	Saving predicted values and residuals when performing an ANCOVA test	213
12.1	SPSS window for univariate ANOVA tests	221
12.2	Saving predicted and residual values when performing a univariate ANOVA test	222
12.3	Applying an ANOVA test to the residuals to test homogeneity of variances	223
12.4	Options pop-up window for requesting homogeneity tests when performing univariate ANOVA tests	223
12.5	Options pop-up window for univariate ANOVA tests	225
12.6	Percentage of morning break spent performing high-intensity activity	227
12.7	Defining a within-within design	228
12.8	Repeated measures of our conceptual dependent variable are assigned to factor level combinations	229
12.9	Saving predicted and residual values when performing an ANOVA test including two within-subjects effects	230
12.10	Selecting options within a repeated measures ANOVA test	230
12.11	Body mass before and after treadmill running in different clothing	233
12.12	Identifying a within-subjects factor within a mixed ANOVA	234
12.13	Defining a mixed ANOVA test	235
12.14	Selecting options in a mixed ANOVA test	236
12.15	Percentage time spent performing high-intensity activity by international and club level netball players in each quarter of a match	239
13.1	Multivariate pop-up window in SPSS	246
13.2	Options pop-up window within the Multivariate facility of SPSS	247
13.3	Post hoc test selection for MANOVA tests	247
13.4	Use of the repeated measures ANOVA facility in SPSS to perform a repeated measures MANOVA	255
13.5	Linking variables (columns) in the SPSS datasheet to repeated measures of conceptual dependent variables	256
14.1	The pop-up window to compare two independent samples using a non-parametric test	270
14.2	Mean rally length (s) in Grand Slam singles tennis	271
14.3	Pop-up window for two related samples tests to be done non-parametrically	273
14.4	Percentage of points won by the serving player in Grand Slam singles tennis	274
14.5	Pop-up window for the Kruskal–Wallis H test	276
14.6	Mean rally duration at Grand Slam tennis tournaments	278
14.7	Pop-up window for the Friedman test in SPSS	280

14.8	Percentage of time spent performing high-intensity activity in netball	282
15.1	Chi Square pop-up window	288
15.2	Birth month distribution of female tennis players participating in Grand Slam singles tennis in 2009	289
15.3	Cross-tabulation pop-up window	292
15.4	Statistics pop-up window for cross-tabulations	292
16.1	The discriminant function analysis pop-up window	302
16.2	Territorial Map (-1 is loss, 0 is draw and +1 is win)	305
16.3	The pop-up window for binary logistic regression	309
17.1	Hierarchical cluster analysis pop-up window	317
17.2	Dendrogram showing the partitioning possibilities for the sample	317
17.3	The four cluster solution	319
17.4	Product ratings by the three different clusters	322
18.1	The Factor Analysis pop-up window	328
18.2	A scree plot	330
19.1	The pivot table cross-tabulating the scores of assessors 'Q1' and 'Q2'	348
19.2	Spreadsheet to determine kappa	349
19.3	Bland-Altman plot for test-retest reliability of the Y-Balance test performed on the dominant leg	355
19.4	Relationship between absolute inter-rater difference and the mean value recorded by raters	356
19.5	Bland-Altman Plot for soccer player valuations	359
19.6	Reliability Analysis pop-up window in SPSS	365
20.1	Power curves for a one-tailed one-sample t-test comparing mean rowing time with a hypothesized upper limit of seven minutes (420s)	373