

DOING BETTER STATISTICS
IN HUMAN-COMPUTER
INTERACTION

Each chapter of this book covers specific topics in statistical analysis, such as robust alternatives to t -tests or how to develop a questionnaire. They also address particular questions on these topics that are commonly asked by HCI researchers when planning or completing the analysis of their data. The book presents the current best practice in statistics, drawing on the state-of-the-art literature that is rarely presented in HCI. This is achieved by providing strong arguments that support good statistical analysis without relying on mathematical explanations. It additionally offers some philosophical underpinning for statistics, so that readers can see how statistics fit with experimental design and the fundamental goal of discovering new HCI knowledge.

PAUL CAIRNS is a Reader in Human-Computer Interaction at the University of York, UK, and Scholar-in-Residence for The AbleGamers Charity, which helps people with disabilities combat social isolation by making videogames more accessible. He has taught statistics at all levels of education for nearly twenty years. His particular research interest is in players' experiences of digital games, and his expertise in experimental and statistical methods was developed through working in this area.

Cambridge University Press
978-1-108-48252-3 — Doing Better Statistics in Human-Computer Interaction
Paul Cairns
Frontmatter
[More Information](#)

DOING BETTER STATISTICS
IN HUMAN–COMPUTER
INTERACTION

PAUL CAIRNS

University of York



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press
978-1-108-48252-3 — Doing Better Statistics in Human-Computer Interaction
Paul Cairns
Frontmatter
[More Information](#)

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom
One Liberty Plaza, 20th Floor, New York, NY 10006, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India
79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.
It furthers the University's mission by disseminating knowledge in the pursuit of
education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108482523
DOI: 10.1017/9781108685139

© Cambridge University Press 2019

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2019

Printed in the United Kingdom by TJ International Ltd, Padstow Cornwall

A catalogue record for this publication is available from the British Library.

ISBN 978-1-108-48252-3 Hardback
ISBN 978-1-108-71059-6 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy
of URLs for external or third-party internet websites referred to in this publication
and does not guarantee that any content on such websites is, or will remain,
accurate or appropriate.

*For HB, my ground, my sky and my sunshine,
and for Samson, whom we miss.*

Cambridge University Press
978-1-108-48252-3 — Doing Better Statistics in Human-Computer Interaction
Paul Cairns
Frontmatter
[More Information](#)

Contents

<i>List of Figures</i>	<i>page xi</i>
<i>List of Tables</i>	xiv
<i>Acknowledgements</i>	xv
Getting Started	I
PART I WHY WE USE STATISTICS	9
1 How Statistics Support Science	11
1.1 The Problem of Induction	12
1.2 Severe Testing	15
1.3 Evidence in HCI	17
1.4 New Experimentalism in HCI	18
1.5 Big Data	21
1.6 Conclusions	22
2 Testing the Null	25
2.1 The Basics of NHST	26
2.2 Going beyond p -Values	30
2.3 NHST and Severe Testing	33
2.4 Honesty in Statistics	35
3 Constraining Bayes	38
3.1 Defining Probability	40
3.2 Plausibility	42
3.3 Unconstrained Bayes	44
3.4 The Bayesian Critique of Frequentism	47
3.5 Being Careful: A Response to the Critique	48
3.6 So, Frequentist or Bayesian?	50
4 Effects: What Tests Test	53
4.1 Location	55
4.2 Dominance	58
4.3 Variation	59
4.4 Estimation and Significance	60

4.5	Big, Small and Zero Effects	62
4.6	Choosing Tests to See Effects	64
PART II HOW TO USE STATISTICS		69
5	Planning Your Statistical Analysis	71
5.1	Principle 1: Articulation	73
5.2	Principle 2: Simplicity	75
5.3	Principle 3: Honesty	77
5.4	Conclusions	79
6	A Cautionary Tail: Why You Should Not Do a One-Tailed Test	80
6.1	A Tale of Two Tails	81
6.2	One-Tail Bad, Two-Tails Better	82
7	Is This Normal?	86
7.1	What Makes Data Normal?	86
7.2	The Problems of Non-normal Data	90
7.3	Testing for Normality	91
7.4	Implications	93
8	Sorting Out Outliers	95
8.1	Detecting Outliers	96
8.2	Sources and Remedies for Outliers	98
8.2.1	Errors in Data	99
8.2.2	Mischievous Participants	99
8.2.3	Faulty Study Design	100
8.2.4	Natural Variation	101
8.3	Conclusions	102
9	Power and Two Types of Error	104
9.1	Type I and Type II Errors	105
9.2	Defining Power	106
9.3	Power and Sample Sizes	108
9.4	Power and the Quality of Tests	110
9.5	Summary	112
10	Using Non-Parametric Tests	114
10.1	The Mechanics of Ranks	115
10.2	Analysing Errors	117
10.2.1	Type I Errors	117
10.2.2	Type II Errors	119
10.3	Practical Use	120
10.4	Reporting Non-Parametric Tests	122
10.5	Summary	123

Contents

ix

11	A Robust t -Test	125
11.1	A Traditional t -Test	126
11.2	Simple Solutions?	130
11.3	Location, Location, Location	131
11.4	Trimmed and Winsorized Means	132
11.5	M-Estimators	134
11.6	Back to t -Tests	135
11.7	Overall Advice	136
12	The ANOVA Family and Friends	139
12.1	What ANOVA Does	140
12.2	Is ANOVA Robust?	145
12.3	Robust Alternatives to ANOVA	147
12.3.1	Non-Parametric Alternatives	147
12.3.2	Changes of Location	149
12.3.3	Do Something Else	150
12.4	Summary	151
13	Exploring, Over-Testing and Fishing	155
13.1	Exploring After a Severe Test	156
13.2	Exploratory Studies	157
13.3	Over-Testing	159
13.3.1	ANOVA Can (Sometimes) Help	160
13.3.2	Planned Comparisons	161
13.3.3	The Bonferroni Correction	161
13.3.4	Bayesian Methods Can Over-Test Too	162
13.4	Fishing	162
13.5	Some Rules of Exploration	164
14	When Is a Correlation Not a Correlation?	167
14.1	Defining Correlation	168
14.2	Outlying Points	170
14.3	Clusters	171
14.4	Avoiding Problems	173
14.5	A Final Warning	174
15	What Makes a Good Likert Item?	175
15.1	Some Important Context	176
15.2	Should Items Have a Midpoint?	178
15.3	How Many Options?	180
15.4	Label All Options or Just End-Points?	181
15.5	The Final Story?	182
16	The Meaning of Factors	186
16.1	From Concepts to Items	188
16.2	From Items to Factors	190

x	<i>Contents</i>	
	16.2.1 The Methods of Factor Analysis	192
	16.2.2 Finding Factors	194
	16.3 From Factors to Concepts?	198
	16.4 What Does It Mean?	200
17	Unreliable Reliability: The Problem of Cronbach's Alpha	204
	17.1 Reliability and Validity	205
	17.2 A Simple Model	206
	17.3 When α Is Low	208
	17.4 When α Is Too High	209
	17.5 Beyond α	211
18	Tests for Questionnaires	214
	18.1 Testing Likert Items	215
	18.1.1 Type I Analysis	216
	18.1.2 Power Analysis	218
	18.1.3 Which Test for Likert Items?	222
	18.2 Questionnaire Data	222
	18.2.1 Type I Analysis	224
	18.2.2 Power Analysis	225
	18.2.3 Which Test for Questionnaires?	229
	18.3 One Final Observation	229
	<i>Index</i>	231

Figures

2.1	A barchart representing the distribution of the responses to a Likert item.	<i>page</i> 29
2.2	A graph representing the normal distribution of the time people take to do a task.	30
4.1	An example of change in location in the normally distributed times of an account entry system.	56
6.1	The t -distribution for 26 degrees of freedom and $t = 1.91$ marked.	82
6.2	The tails in the t -distribution with 26 degrees of freedom.	83
7.1	The classic bell curve of a normal distribution with mean 0 and standard deviation of 1. The shaded area represents the data within one standard deviation of the mean and corresponds to about 67% of the area under the curve.	87
7.2	A binomial distribution for a fair coin flipped 20 times and its normal approximation.	88
7.3	A histogram of the means of 10,000 samples of size 20 taken from a lognormal distribution and the normal curve that a parametric test would use to approximate the histogram.	90
8.1	The boxplot of random samples of size 29 from each of a normal, lognormal and χ^2 distribution. Outliers are indicated as circles, though because outliers are defined with reference to the normal distribution, these points are not in fact outliers for the distributions they belong to.	98
10.1	Two distributions that differ only in variance but nonetheless affect the Type I error rate of the Mann–Whitney test.	118
10.2	Two distributions that differ in dominance but where one also spreads out across the other making a Mann–Whitney test low in power.	121

xii	<i>List of Figures</i>	
11.1	The probability distribution curve of a normal distribution with standard deviation of 1 and a mixed-normal curve with a standard deviation of 3.3.	128
11.2	A polarised distribution on a Likert scale: 100 people are equally divided between those who strongly agree and those who strongly disagree.	132
12.1	Hypothetical interaction plots for one effect in a two-way ANOVA when both other effects are zero.	143
14.1	Scatterplots with regression lines demonstrating three different correlation coefficients between two normally distributed variables.	169
14.2	The weakening and strengthening effect of a single outlier on correlation coefficients of 50 other data points. In each case, the outlier is the leftmost point.	171
14.3	The effect of clustering on correlation: the correlation within the left cluster is $r = 0.088$, within the right cluster is $r = 0.17$ and of the two groups together is $r = 0.30$.	172
15.1	A surprisingly current example item from Likert's 1932 attitude questionnaire.	176
15.2	Three formats of Likert item: 1. with midpoint; 2. no midpoint; 3. no midpoint but option to have no opinion	178
15.3	Two formats of Likert item: 1. fully labelled; 2. end-points labelled.	182
16.1	Scree plot of the competitive social presence component of the CCPIG	195
17.1	An example model for Ensmartness with six items. All items are correlated to a true Ensmartness value, T , and half to the unmeasured factor, U .	207
18.1	Boxplots of number of significant Type I errors made by each test over 10,000 trials for each Likert distribution and for different sample sizes. The dashed lines are at 500 tests corresponding to $\alpha = 0.05$.	218
18.2	Boxplots of number of significant trials out of 10,000 trials for each pair of distinct Likert distributions and for two different sample sizes.	219
18.3	Scatterplots of highest power test against the ideal difference in means or dominance between the distributions.	221
18.4	The varying density distributions of four of the datasets used in the analysis.	225

<i>List of Figures</i>		xiii
18.5	Boxplots of number of significant Type I errors made by each test over 10,000 trials for each questionnaire distribution and for different sample sizes. The dashed line is at 500 tests corresponding to $\alpha = 0.05$.	226
18.6	Boxplots of number of significant trials out of 10,000 trials for each pair of distinct questionnaire distributions and for two different sample sizes.	227
18.7	Scatterplots of highest power test against the ideal difference in means or dominance between the distributions.	228

Tables

9.1	The relationship of a test decision to the real world.	<i>page</i> 105
12.1	The different sources of variation in a one-way between-participants ANOVA where X_i represents the data point gathered from a single participant.	142
16.1	Loadings of items on factors for different factor solutions	196
16.2	Loadings of items on the one-factor and bifactor models of competitive social presence with two specific factors, Eng = Engagement and Aware = Awareness.	199
17.1	A simplification of the dissociation between validity and reliability of a scale and the quality of the resulting scale	205
18.1	The different distributions (as percentage probabilities) of 5-point Likert item data proposed by De Winter and Dodou (2010).	216
18.2	Summary details of the sixteen scales of realistic questionnaire data.	224

Acknowledgements

This book is a culmination of a lot of thinking, teaching and thinking about teaching around statistics and accordingly there are lots of people who have had a hand in getting me to this point.

I would like to thank all of my PhD students, at York and at Goldsmiths, for their encouragement and enthusiasm for this project all the while I have been on sabbatical. I would like particularly to thank Joe Cutting for his concrete and constructive feedback on early versions of chapters. I would also like to thank all those research students explicitly mentioned in the book for giving me such rich, diverse and interesting research (and resulting data) to think about.

Many thanks go to my good colleagues Helen Petrie and Chris Power. Much of this book was written while on sabbatical and they took on my teaching load to make the sabbatical possible. Also, I would like to thank Chris further for his patience in listening to me enthuse about statistics and also forcing me to explain things properly. Thanks too to my department for allowing me to take the sabbatical.

Jan Böhnke has also been a helpful support over the years as I have grown my statistical knowledge. He has challenged my assumptions through many valuable discussions. Harold Thimbleby has been a great support throughout my academic career and his feedback on this book was the just right thing at just the right time.

I would like to thank my many students at all levels of higher education. They put up with me going on about statistics a lot but it is through my attempts, successful and unsuccessful, at trying to teach them statistics that I have gained a lot of the experiences that inform this book.

Finally, I would like to thank my family. Most of my family have been encouraging even in the face of what must have seemed like glacial progress in writing. Thank you for your support and belief. Two key members of my family who have been less supportive are my children, Eleanor and Patrick.

Their cheerful lack of interest in my work has been an important grounding both for this book and life generally. And last but not most of all, thanks to HB for being my constant companion and friend and, also, for giving useful feedback on this book as it progressed.