

Cambridge University Press

052183550X - Statistics Explained: An Introductory Guide for Life Scientists

Steve McKillup

Frontmatter

[More information](#)

Statistics Explained

Statistics Explained is a reader-friendly introduction to experimental design and statistics for undergraduate students in the life sciences, particularly those who do not have a strong mathematical background. Hypothesis testing and experimental design are discussed first. Statistical tests are then explained using pictorial examples and a minimum of formulae. This class-tested approach, along with a well-structured set of diagnostic tables, will give students the confidence to choose an appropriate test with which to analyse their own data sets. Presented in a lively and straightforward manner *Statistics Explained* will give readers the depth and background necessary to proceed to more advanced texts and applications. It will therefore be essential reading for all bioscience undergraduates, and will serve as a useful refresher course for more advanced students.

Steve McKillup is an Associate Professor of Biology in the School of Biological and Environmental Sciences at Central Queensland University, Rockhampton.

Cambridge University Press

052183550X - Statistics Explained: An Introductory Guide for Life Scientists

Steve McKillup

Frontmatter

[More information](#)

Statistics Explained

An Introductory Guide for Life Scientists

STEVE McKILLUP



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press

052183550X - Statistics Explained: An Introductory Guide for Life Scientists

Steve McKillup

Frontmatter

[More information](#)

CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

CAMBRIDGE UNIVERSITY PRESS

The Edinburgh Building, Cambridge CB2 2RU, UK

www.cambridge.org

Information on this title: www.cambridge.org/9780521835503

© S. McKillup 2005

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 2005

Printed in the United Kingdom at the University Press, Cambridge

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

ISBN-13 978-0-521-83550-3 hardback

ISBN-10 0-521-83550-X hardback

ISBN-13 978-0-521-54316-3 paperback

ISBN-10 0-521-54316-9 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.

Contents

	<i>Preface</i>	<i>page xi</i>
1	Introduction	1
1.1	Why do life scientists need to know about experimental design and statistics?	1
1.2	What is this book designed to do?	5
2	‘Doing science’ – hypotheses, experiments, and disproof	7
2.1	Introduction	7
2.2	Basic scientific method	7
2.3	Making a decision about an hypothesis	10
2.4	Why can’t an hypothesis or theory ever be proven?	11
2.5	‘Negative’ outcomes	11
2.6	Null and alternate hypotheses	12
2.7	Conclusion	13
3	Collecting and displaying data	14
3.1	Introduction	14
3.2	Variables, experimental units, and types of data	14
3.3	Displaying data	16
3.4	Displaying ordinal or nominal scale data	20
3.5	Bivariate data	23
3.6	Multivariate data	25
3.7	Summary and conclusion	26
4	Introductory concepts of experimental design	27
4.1	Introduction	27
4.2	Sampling – mensurative experiments	28

vi	Contents	
4.3	Manipulative experiments	32
4.4	Sometimes you can only do an unreplicated experiment	39
4.5	Realism	40
4.6	A bit of common sense	41
4.7	Designing a ‘good’ experiment	41
4.8	Conclusion	42
5	Probability helps you make a decision about your results	44
5.1	Introduction	44
5.2	Statistical tests and significance levels	45
5.3	What has this got to do with making a decision or statistical testing?	49
5.4	Making the wrong decision	49
5.5	Other probability levels	50
5.6	How are probability values reported?	51
5.7	All statistical tests do the same basic thing	52
5.8	A very simple example – the chi-square test for goodness of fit	52
5.9	What if you get a statistic with a probability of exactly 0.05?	55
5.10	Statistical significance and biological significance	55
5.11	Summary and conclusion	55
6	Working from samples – data, populations, and statistics	57
6.1	Using a sample to infer the characteristics of a population	57
6.2	Statistical tests	57
6.3	The normal distribution	57
6.4	Samples and populations	63
6.5	Your sample mean may not be an accurate estimate of the population mean	65
6.6	What do you do when you only have data from one sample?	67
6.7	Why are the statistics that describe the normal distribution so important?	71
6.8	Distributions that are not normal	72
6.9	Other distributions	73
6.10	Other statistics that describe a distribution	74
6.11	Conclusion	75

	Contents	vii
7	Normal distributions – tests for comparing the means of one and two samples	77
7.1	Introduction	77
7.2	The 95% confidence interval and 95% confidence limits	77
7.3	Using the Z statistic to compare a sample mean and population mean when population statistics are known	78
7.4	Comparing a sample mean with an expected value	81
7.5	Comparing the means of two related samples	88
7.6	Comparing the means of two independent samples	90
7.7	Are your data appropriate for a <i>t</i> test?	92
7.8	Distinguishing between data that should be analysed by a paired sample test or a test for two independent samples	94
7.9	Conclusion	95
8	Type 1 and Type 2 errors, power, and sample size	96
8.1	Introduction	96
8.2	Type 1 error	96
8.3	Type 2 error	97
8.4	The power of a test	100
8.5	What sample size do you need to ensure the risk of Type 2 error is not too high?	102
8.6	Type 1 error, Type 2 error, and the concept of biological risk	104
8.7	Conclusion	104
9	Single factor analysis of variance	105
9.1	Introduction	105
9.2	Single factor analysis of variance	106
9.3	An arithmetic/pictorial example	112
9.4	Unequal sample sizes (unbalanced designs)	117
9.5	An ANOVA does not tell you which particular treatments appear to be from different populations	117
9.6	Fixed or random effects	118
10	Multiple comparisons after ANOVA	119
10.1	Introduction	119
10.2	Multiple comparison tests after a Model I ANOVA	119

viii	Contents	
10.3	An a-posteriori Tukey comparison following a significant result for a single factor Model I ANOVA	122
10.4	Other a-posteriori multiple comparison tests	123
10.5	Planned comparisons	124
11	Two factor analysis of variance	127
11.1	Introduction	127
11.2	What does a two factor ANOVA do?	129
11.3	How does a two factor ANOVA analyse these data?	131
11.4	How does a two factor ANOVA separate out the effects of each factor and interaction?	136
11.5	An example of a two factor analysis of variance	139
11.6	Some essential cautions and important complications	140
11.7	Unbalanced designs	149
11.8	More complex designs	149
12	Important assumptions of analysis of variance: transformations and a test for equality of variances	151
12.1	Introduction	151
12.2	Homogeneity of variances	151
12.3	Normally distributed data	152
12.4	Independence	155
12.5	Transformations	156
12.6	Are transformations legitimate?	158
12.7	Tests for heteroscedasticity	159
13	Two factor analysis of variance without replication, and nested analysis of variance	162
13.1	Introduction	162
13.2	Two factor ANOVA without replication	162
13.3	A-posteriori comparison of means after a two factor ANOVA without replication	166
13.4	Randomised blocks	167
13.5	Nested ANOVA as a special case of a one factor ANOVA	168
13.6	A pictorial explanation of a nested ANOVA	170
13.7	A final comment on ANOVA – this book is only an introduction	175

	Contents	ix
14	Relationships between variables: linear correlation and linear regression	176
14.1	Introduction	176
14.2	Correlation contrasted with regression	177
14.3	Linear correlation	177
14.4	Calculation of the Pearson r statistic	178
14.5	Is the value of r statistically significant?	184
14.6	Assumptions of linear correlation	184
14.7	Summary and conclusion	184
15	Simple linear regression	186
15.1	Introduction	186
15.2	Linear regression	186
15.3	Calculation of the slope of the regression line	188
15.4	Calculation of the intercept with the Y axis	192
15.5	Testing the significance of the slope and the intercept of the regression line	193
15.6	An example – mites that live in the your hair follicles	199
15.7	Predicting a value of Y from a value of X	201
15.8	Predicting a value of X from a value of Y	201
15.9	The danger of extrapolating beyond the range of data available	202
15.10	Assumptions of linear regression analysis	202
15.11	Further topics in regression	204
16	Non-parametric statistics	205
16.1	Introduction	205
16.2	The danger of assuming normality when a population is grossly non-normal	205
16.3	The value of making a preliminary inspection of the data	207
17	Non-parametric tests for nominal scale data	208
17.1	Introduction	208
17.2	Comparing observed and expected frequencies – the chi-square test for goodness of fit	209
17.3	Comparing proportions among two or more independent samples	212

x	Contents	
17.4	Bias when there is one degree of freedom	215
17.5	Three-dimensional contingency tables	219
17.6	Inappropriate use of tests for goodness of fit and heterogeneity	220
17.7	Recommended tests for categorical data	221
17.8	Comparing proportions among two or more related samples of nominal scale data	222
18	Non-parametric tests for ratio, interval, or ordinal scale data	224
18.1	Introduction	224
18.2	A non-parametric comparison between one sample and an expected distribution	225
18.3	Non-parametric comparisons between two independent samples	227
18.4	Non-parametric comparisons among more than two independent samples	232
18.5	Non-parametric comparisons of two related samples	236
18.6	Non-parametric comparisons among three or more related samples	238
18.7	Analysing ratio, interval, or ordinal data that show gross differences in variance among treatments and cannot be satisfactorily transformed	241
18.8	Non-parametric correlation analysis	243
18.9	Other non-parametric tests	245
19	Choosing a test	246
19.1	Introduction	246
20	Doing science responsibly and ethically	255
20.1	Introduction	255
20.2	Dealing fairly with other people's work	255
20.3	Doing the experiment	257
20.4	Evaluating and reporting results	258
20.5	Quality control in science	260
	<i>References</i>	261
	<i>Index</i>	263

Cambridge University Press

052183550X - Statistics Explained: An Introductory Guide for Life Scientists

Steve McKillup

Frontmatter

[More information](#)

Preface

If you mention ‘statistics’ or ‘biostatistics’ to life scientists, they often look nervous. Many fear or dislike mathematics, but an understanding of statistics and experimental design is essential for graduates, postgraduates, and researchers in the biological, biochemical, health, and human movement sciences.

Since this understanding is so important, life science students are usually made to take some compulsory undergraduate statistics courses. Nevertheless, I found that a lot of graduates (and postgraduates) were unsure about designing experiments and had difficulty knowing which statistical test to use (and which ones not to!) when analysing their results. Some even told me they had found statistics courses ‘boring, irrelevant and hard to understand’.

It seemed there was a problem with the way many introductory biostatistics courses were presented, which was making students disinterested and preventing them from understanding the concepts needed to progress to higher-level courses and more complex statistical applications. There seemed to be two major reasons for this problem, and as a student I encountered both.

First, a lot of statistics textbooks take a mathematical approach and often launch into considerable detail and pages of daunting looking formulae without any straightforward explanation about what statistical testing really does.

Second, introductory biostatistics courses are often taught in a way that does not cater for life science students who may lack a strong mathematical background.

When I started teaching at Central Queensland University I thought there had to be a better way of introducing essential concepts of

Cambridge University Press

052183550X - Statistics Explained: An Introductory Guide for Life Scientists

Steve McKillup

Frontmatter

[More information](#)

xii Preface

biostatistics and experimental design. It had to start from first principles and develop an understanding that could be applied to all statistical tests. It had to demystify what these tests actually did and explain them with a minimum of formulae and terminology. It had to relate statistical concepts to experimental design. And, finally, it had to build a strong understanding to help the student progress to more complex material. I tried this approach with my undergraduate classes and the response from a lot of students, including some postgraduates who sat in on the course, was, ‘Hey Steve, you should write an introductory stats book!’

Ward Cooper suggested I submit a proposal for this sort of book to Cambridge University Press. Ruth McKillup read, commented on, and reread several drafts, provided constant encouragement, and tolerated my absent mindedness. My students, especially Steve Dunbar, Kevin Strychar, and Glenn Druery, encouraged me to start writing and my friends and colleagues, especially Dearne Mayer and Sandy Dalton, encouraged me to finish. Finally, I sincerely thank the anonymous reviewers of the initial proposal and the subsequent manuscript who, without exception, made most appropriate suggestions for improvement.