

Preface to the second edition

I have been astounded by the positive reception the first edition of my little book has received. It has clearly filled a need; students from many different countries have emailed me just to say 'Thank you for writing the book.' Student reviews on Amazon have given my book high praise, and colleagues teaching statistics to biologists have also found it of value. One review complained about the 'lack of maths'. I take this as a compliment, and the review arguing that the availability of computer packages makes it unnecessary to be able to carry out the calculations 'long-hand' seems to me to completely miss the point!

But I have to hang my head in shame at the numerous errors there were in the first printing of 2008. Particularly the numerical examples were littered with computational errors (not infrequently in the form of dyslexic transpositions). The book went through several iterations before going to press, and I really should have rechecked all the calculations in the final proof. I can only apologise. However, these many errors led to a very encouraging outcome. The errors were largely identified by users, who then contacted me convinced they had found an error. The statistics teaching I received never gave me similar confidence. That students were sure they were right and that I had made an error is almost the best evidence of the success of my book that I could have asked for.

So what's new in this second edition? The statistics that my book tries to teach are the concepts elaborated by R.A. Fisher in the 1930s, and today these remain the basis of the statistical procedures based on the normal distribution. I can tweak the English, change sections to make them more easily understood, and add some further extensions of Fisher's concepts – and all this I have done. I have been surprised how I, as a reader of the text of the first edition, have found sentences completely incomprehensible that were written by me as author 10 years ago! I have used the opportunity of the second edition to try to clarify these passages for both you and me!

I have greatly revised and, I hope, simplified the chapter on linear regression and correlation and added some material to the chapter on chi-square tests. However, the most significant addition has been that analysis of covariance, just briefly mentioned in the first edition, has been allocated a chapter of its own. It deserves emphasis as a valuable statistical technique, and I find most textbooks shroud it in mystery. Although I would advise you to use computer programs if you want to use analysis of covariance, I thought it a good idea to demystify it as much as I can. The calculation of numbers in analysis of covariance is usually presented as having no connection with more familiar calculations, yet they come from the standard techniques of analysis of variance and regression. The point is that only some of the resulting values are then used while the rest are discarded, but doing the analyses that include these redundant numbers does make the whole thing much more comprehensible.

I can only hope that this second edition proves as popular as the first.

Helmut F. van Emden
July 2018

Preface to the first edition

I have written/edited several books on my own speciality, agricultural and horticultural entomology, but always at the request of a publisher or colleagues. This book is different in two ways. Firstly, it is a book I have positively wanted to write for many years and secondly, I am stepping out of my 'comfort zone' in doing so.

The origins of this book stem from my appointment to the Horticulture Department of Reading University under Professor O. V. S. Heath, FRS. Professor Heath appreciated the importance of statistics and, at a time when there was no University-wide provision of statistics teaching, he taught the final year students himself. I became the 'assistant' whose role it was to run the practical exercises which followed the Professor's lecture. You cannot teach what you do not understand yourself, but I tried nonetheless.

Eventually I took over the entire course. By then it was taught in the second year and in the third year the students went on to take a Faculty-wide course. I did not continue the lectures; the whole course was in the laboratory where I led the students (using pocket calculators) through the calculations in stages. The laboratory class environment involved continuous interaction with students in a way totally different from what happens in lectures, and it rapidly became clear to me that many biologists have their neurons wired up in a way that makes the traditional way of teaching statistics rather difficult for them.

What my students needed was confidence – confidence that statistical ideas and methods were not just theory, but actually worked with real biological data and, above all, had some basis in logic! As the years of teaching went on, I began to realise that the students regularly found the same steps a barrier to progress and a damage to their confidence. Year after year I tried new ways to help them over these 'crisis points'; eventually I succeeded with all of them, I am told.

The efficacy of my unusual teaching aids can actually be quantified. After the Faculty course taught by professional statisticians, my students were formally examined together with cohorts of students from other departments in the Faculty (then of 'Agriculture and Food') who had attended the same third year course in Applied Statistics. My students mostly (perhaps all but three per year out of some 20) appeared in the mark list as a distinct block right at the upper end, with percentage marks in the 70s, 80s and even 90s. Although there may have also been one or two from other courses with high marks, there then tended to be a gap till marks in the lower 60s appeared and began a continuum down to single figures.

I therefore feel confident that this book will be helpful to biologists with its mnemonics such as SqADS and 'you go along the corridor before you go upstairs' Other things previously unheard of are the 'lead line' and 'super-totals' with their 'subscripts' – yet all have been appreciated as most helpful by my students over the years. A riffle through the pages will amaze – where are the equations and algebraic symbols? They have to a large extent been replaced by numbers and words. The biologists I taught – and I don't think they were atypical – could work out what to do with a '45', but rarely what to do with an 'x'. Also, I have found that there are a number of statistical principles students easily forget, and then inevitably run into trouble with their calculations. These concepts are marked with the symbol of a small

elephant



The book limits itself to the basic foundations of parametric statistics, the t-test, analysis of variance, linear regression and chi-square. However, the reader is guided as to where there are important extensions of these techniques, and there is an introduction to non-parametric tests which includes a check list of non-parametric methods linked to their parametric counterparts. Many chapters end with an 'executive summary' as a quick source for revision, and there are additional exercises to give the practice which is so essential to learning.

In order to minimise algebra, the calculations are explained with numerical examples. These, as well as the 'spare-time activity' exercises have come from many sources, and I regret the origin of many has become lost in the mists of time. Quite a number come from experiments carried out by Horticulture students at Reading as part of their second year outdoor practicals, and others have been totally fabricated in order to 'work out' well. Others have had numbers or treatments changed better to fit what was needed. I can only apologise to anyone whose data I have used without due acknowledgement; failure to do so is not intentional. But please remember that data

have often been fabricated or massaged – therefore do not rely on the results as scientific evidence for what they appear to show!

Today, computer programmes take most of the sweat out of statistical procedures, and most biologists have access to professional statisticians. ‘Why bother to learn basic statistics?’ is therefore a perfectly fair question, akin to ‘Why keep a dog and bark?’ The question deserves an answer; to save repetition, my answer can be found towards the end of Chapter 1.

I am immensely grateful to the generations of Reading students who have challenged me to overcome their ‘hang-ups’ and who have therefore contributed substantially to any success this book achieves. Also many postgraduate students as well as experienced visiting overseas scientists have encouraged me to turn my course into book form. My love and special thanks go to my wife Gillian who, with her own experience of biological statistics, has supported and encouraged me in writing this book; it is to her that I owe the imaginative title for the book.

Finally, I should like to thank Ward Cooper of Blackwells for having faith in this biologist, who is less terrified of statistics than he once was.

Helmut F. van Emden
December 2006