

Preface

The subject of regression, or of the linear model, is central to the subject of statistics. It concerns what can be said about some quantity of interest, which we may not be able to measure, starting from information about one or more other quantities, in which we may not be interested but which we can measure. We model our variable of interest as a linear combination of these variables (called covariates), together with some error. It turns out that this simple prescription is very flexible, very powerful and useful.

If only because regression is inherently a subject in two or more dimensions, it is not the first topic one studies in statistics. So this book should not be the first book in statistics that the student uses. That said, the statistical prerequisites we assume are modest, and will be covered by any first course on the subject: ideas of sample, population, variation and randomness; the basics of parameter estimation, hypothesis testing, p -values, confidence intervals etc.; the standard distributions and their uses (normal, Student t , Fisher F and chi-square – though we develop what we need of F and chi-square for ourselves).

Just as important as a first course in statistics is a first course in probability. Again, we need nothing beyond what is met in any first course on the subject: random variables; probability distribution and densities; standard examples of distributions; means, variances and moments; some prior exposure to moment-generating functions and/or characteristic functions is useful but not essential (we include all we need here). Our needs are well served by John Haigh's book *Probability models* in the SUMS series, Haigh (2002).

Since the terms regression and linear model are largely synonymous in statistics, it is hardly surprising that we make extensive use of linear algebra and matrix theory. Again, our needs are well served within the SUMS series, in the two books by Blyth and Robertson, *Basic linear algebra* and *Further linear algebra*, Blyth and Robertson (2002a), (2002b). We make particular use of the

material developed there on sums of orthogonal projections. It will be a pleasure for those familiar with this very attractive material from pure mathematics to see it being put to good use in statistics.

Practical implementation of much of the material of this book requires computer assistance – that is, access to one of the many specialist statistical packages. Since we assume that the student has already taken a first course in statistics, for which this is also true, it is reasonable for us to assume here too that the student has some prior knowledge of and experience with a statistical package. As with any other modern student text on statistics, one is here faced with various choices. One does not want to tie the exposition too tightly to any one package; one cannot cover all packages, and shouldn't try – but one wants to include some specifics, to give the text focus. We have relied here mainly on S-Plus/R®.¹

Most of the contents are standard undergraduate material. The boundary between higher-level undergraduate courses and Master's level courses is not a sharp one, and this is reflected in our style of treatment. We have generally included complete proofs except in the last two chapters on more advanced material: Chapter 8, on Generalised Linear Models (GLMs), and Chapter 9, on special topics. One subject going well beyond what we cover – Time Series, with its extensive use of autoregressive models – is commonly taught at both undergraduate and Master's level in the UK. We have included in the last chapter some material, on non-parametric regression, which – while no harder – is perhaps as yet more commonly taught at Master's level in the UK.

In accordance with the very sensible SUMS policy, we have included exercises at the end of each chapter (except the last), as well as worked examples. One then has to choose between making the book more student-friendly, by including solutions, or more lecturer-friendly, by not doing so. We have nailed our colours firmly to the mast here by including full solutions to all exercises. We hope that the book will nevertheless be useful to lecturers also (e.g., in inclusion of references and historical background).

Rather than numbering equations, we have labelled important equations acronymically (thus the normal equations are (*NE*), etc.), and included such equation labels in the index. Within proofs, we have occasionally used local numbering of equations: (*), (*a*), (*b*) etc.

In pure mathematics, it is generally agreed that the two most attractive subjects, at least at student level, are complex analysis and linear algebra. In statistics, it is likewise generally agreed that the most attractive part of the subject is

¹ S+, S-PLUS, S+FinMetrics, S+EnvironmentalStats, S+SeqTrial, S+SpatialStats, S+Wavelets, S+ArrayAnalyzer, S-PLUS Graphlets, Graphlet, Trellis, and Trellis Graphics are either trademarks or registered trademarks of Insightful Corporation in the United States and/or other countries. Insightful Corporation 1700 Westlake Avenue N, Suite 500 Seattle, Washington 98109 USA.

regression and the linear model. It is also extremely useful. This lovely combination of good mathematics and practical usefulness provides a counter-example, we feel, to the opinion of one of our distinguished colleagues. Mathematical statistics, Professor x opines, combines the worst aspects of mathematics with the worst aspects of statistics. We profoundly disagree, and we hope that the reader will disagree too.

The book has been influenced by our experience of learning this material, and teaching it, at a number of universities over many years, in particular by the first author's thirty years in the University of London and by the time both authors spent at the University of Sheffield. It is a pleasure to thank Charles Goldie and John Haigh for their very careful reading of the manuscript, and Karen Borthwick and her colleagues at Springer for their kind help throughout this project. We thank our families for their support and forbearance.

NHB, JMF

Imperial College, London and the University of East London, March 2010

Regression

Linear Models in Statistics

Bingham, N.H.; Fry, J.M.

2010, XIII, 284 p. 50 illus., Softcover

ISBN: 978-1-84882-968-8