Tracking data can be baffling without a thorough knowledge of statistical approaches.

STATISTICS

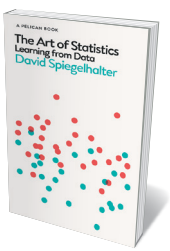
The joy of stats

Evelyn Lamb enjoys a rich study on number-crunching and its ubiquitous fruit.

The aphorism "All models are wrong, but some are useful", attributed to statistician George Box, is a cliché for a reason. It cuts to the heart of a central challenge facing researchers in many areas of science. The world is more complicated than anything that a mathematical, scientific or statistical model can capture. Yet models of the world, however imperfect, are necessary for drawing conclusions about everything from pharmaceutical efficacy to unemployment numbers. David Spiegelhalter's *The Art of Statistics* shines a light on how we can use the ever-growing deluge of data to improve our understanding of the world — and of some of the pitfalls we encounter in the attempt.

The book is part of a trend in statistics education towards emphasizing conceptual understanding rather than computational fluency. Statistics software can now perform a battery of tests and crunch any measure from large data sets in the blink of an eye. Thus, being able to compute the standard deviation of a sample the long way is seen as less essential than understanding how to design and interpret scientific studies with a rigorous eye.

Throughout the book, Spiegelhalter emphasizes the importance of the "PPDAC" structure: Problem-Plan-Data-Analysis-Conclusion. He describes how statisticians approach each section of an investigation, and the tools that come into play. PPDAC starts with defining a problem or question and developing a plan for what to measure, how to measure it and what analyses will serve best. Then researchers collect data, analyse them according to the plan and decide what conclusions reasonably follow.



The Art of Statistics: Learning from Data
DAVID SPIEGELHALTER
Pelican (2019)

Spiegelhalter has had a long and prominent career as a statistician, and some of the most interesting passages occur when he pulls back the curtain and describes the choices he and his colleagues have made in the course of research. For example, he led a team that investigated death rates in children after heart surgery in British hospitals. Even for seemingly unambiguous figures — such as the number of children who had heart surgery and how many died — there were discrepancies between data sources and definitions.

Decisions about how to treat potential ambiguities and edge cases can have a large effect on the outcome of a study. For instance, for how long after an operation should a death be attributable to it? Because complete clarity and objectivity are not possible, people who read about these studies — especially politicians or jurors who have to decide whether unusual numbers of deaths owing to surgery point to crimes or malpractice — should be aware of the potential complicating factors, if only to temper their confidence in a finding.

As such examples show, a main takeaway from the book is a sense of circumspection about our confidence in what is known. As

SOLARSEVEN/GETTY

© 2019 Springer Nature Limited. All rights reserved.

Spiegelhalter writes, the point of statistical science is to ease us through the stages of extrapolation from a controlled study to an understanding of the real world, "and finally, with due humility, be able to say what we can and cannot learn from data". That humility can be lacking when statistics are used in debates about contentious issues such as the costs and benefits of cancer screening.

PACKED SYLLABUS

The book does an admirable job of covering a great deal of ground in limited space. Some concepts would have benefited from a deeper treatment: notably, bootstrapping, or estimating the distribution of a statistic on the basis of random resampling; and the central limit theorem, which holds that averages of increasingly large subsets of the data in many sets tend towards a normal distribution. However, Spiegelhalter had difficult decisions to make about how much of each topic he would unpack. A book covering the ideas of regression, null-hypothesis testing, Bayesian inference and much, much more cannot be comprehensive. The robust notes and bibliography will be useful for readers who wish to delve deeper.

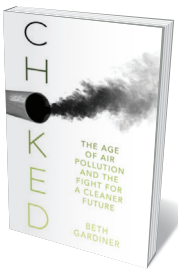
Spiegelhalter does not shy away from discussions of subtle statistical issues such as the nature of different types of uncertainty. So, as he warns at the beginning of chapter 9, where the rubber of mathematical probability theory hits the road of statistical inference, some material will prove challenging even to scientifically sophisticated readers. Some passages require pencil, paper and a few passes through to fully digest, but the approachable big-picture explanations and end-of-chapter summaries help, as does the glossary.

A useful coda focuses on the many dubious statistical practices that have helped to create today's replication crisis across swathes of science. Spiegelhalter touches, too, on how both scientists and journalists can improve public understanding by running better studies and reporting on them responsibly. After wading into the statistical depths earlier in the book, readers can start using these tangible, easily applicable lessons immediately.

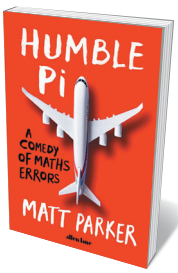
The Art of Statistics will serve students well. And it will be a boon for journalists eager to use statistics responsibly — along with anyone who wants to approach research and its reportage with healthy scepticism. ■

Evelyn Lamb *is a freelance math and science writer in Salt Lake City, Utah. Her work has appeared in publications including* Scientific American, Slate, Nautilus *and* Quanta.
e-mail: rootsofunityblog@gmail.com

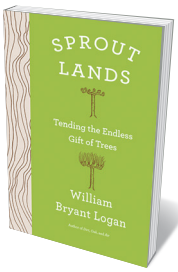
Books in brief



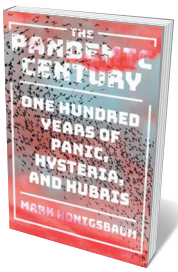
Choked
Beth Gardiner GRANTA *(2019)*
Take a breath. Chances are, you've just inhaled a foul airborne soup — the toxic particulates, nitrogen oxides and more emanating from fossil-fuel combustion. Atmospheric pollution now kills 7 million people a year. To pierce the murk around this issue, journalist Beth Gardiner travelled through Delhi (smog's 'ground zero') to hotspots in the United Kingdom, Poland, Malawi and beyond. She met researchers transforming understanding, rickshaw drivers in the firing line and policy experts struggling with the vagaries of regulation in the United States and China. Timely, eloquent and disturbing.



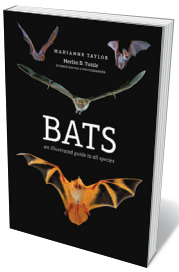
Humble Pi
Matt Parker ALLEN LANE *(2019)*
We live in a world supersaturated with mathematics — and errors are rife, reveals maths writer and stand-up comic Matt Parker in this zinger of a book. Parker takes us through cases in finance, engineering and information technology. He recounts the moment in 2004 when mayhem hit Californian airspace because of a maths-related computer glitch; explains why a genetics textbook retailed at more than US$23 million on Amazon; and exposes oddities in postcodes, units, lotteries, bridge design, spreadsheets and much more. All in all, a hectic ride ever teetering between horror and hilarity.



Sprout Lands
William Bryant Logan W. W. NORTON *(2019)*
"Live wood just won't quit" — so declares arborist and writer William Bryant Logan in this vividly insightful exploration of tree regeneration. Logan documents how roots and stems sprout from myriad points on the body of trees; hymns the joys of hedge-laying; and examines the pollarding and coppicing legacies of Neolithic Europeans, sixteenth-century Basques and California's Indigenous Karuk people. Surveying New York City's Fresh Kills dump, he shows how the journeywork of red maples and mulberries returns our detritus to nature and offers a metaphor for hope in a changing world.



The Pandemic Century
Mark Honigsbaum W. W. NORTON *(2019)*
Over the past 100 years, argues medical historian Mark Honigsbaum, scientific paradigms have often blinded researchers to coming pandemics. Thus, in 1918, the bacteriologist Richard Pfeiffer's belief that influenza was caused by a bacillus misdirected research even as Spanish flu ravaged millions. Honigsbaum's gripping narrative ranges from the psittacosis pandemic of the late 1920s to the Ebola crisis of 2014. But despite advances in fields such as immunology, we remain poor at predicting outbreaks and often fail to control panic, or to factor in environmental context.



Bats: An Illustrated Guide to All Species
Marianne Taylor and Merlin D. Tuttle IVY *(2019)*
Active in darkness, aflit in a "three-dimensional maze of sound", bats are both enigmatic and, with more than 1,300 species, astoundingly diverse. This guide by writer Marianne Taylor and bat conservationist Merlin Tuttle shines a light on the order Chiroptera, from the wee Kitti's hog-nosed bat (*Craseonycteris thonglongyai*, a candidate for world's smallest mammal) to the 'megabats' of the Pteropodidae family. Meshing deft scientific text with Tuttle's sumptuous images, it's a superb introduction to the baroque morphologies and flying prowess of these beguiling beasts. **Barbara Kiser**

© 2019 Springer Nature Limited. All rights reserved.