# *Using Statistical Methods for Water Quality Management*

## *Issues, Problems and Solutions*

This Page Intentionally Left Blank

*Using Statistical Methods for Water Quality Management*

WILEY SERIES IN STATISTICS IN PRACTICE

Advisory Editor, MARIAN SCOTT, *University of Glasgow, Scotland, UK*

Founding Editor, VIC BARNETT, *Nottingham Trent University, UK*

*Statistics in Practice* is an important international series of texts which provide detailed coverage of statistical concepts, methods and worked case studies in specific fields of investigation and study.

With sound motivation and many worked practical examples, the books show in down-to-earth terms how to select and use an appropriate range of statistical techniques in a particular practical field within each title's special topic area.

The books provide statistical support for professionals and research workers across a range of employment fields and research environments. Subject areas covered include medicine and pharmaceutics; industry, finance and commerce; public services; the earth and environmental sciences, and so on.

The books also provide support to students studying statistical courses applied to the above areas. The demand for graduates to be equipped for the work environment has led to such courses becoming increasingly prevalent at universities and colleges.

It is our aim to present judiciously chosen and well-written workbooks to meet everyday practical needs. Feedback of views from readers will be most valuable to monitor the success of this aim.

A complete list of titles in this series appears at the end of the volume.

# *Using Statistical Methods for Water Quality Management*

## *Issues, Problems and Solutions*

Graham B. McBride
National Institute of Water & Atmospheric Research
Hamilton, New Zealand

WILEY-INTERSCIENCE
A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.



***Library of Congress Cataloging-in-Publication Data:***

McBride, Graham B., 1948–
Using statistical methods for water quality management : issues, problems and solutions / Graham B. McBride.
p. cm.
Includes bibliographical references and index.
ISBN 0-471-47016-3 (cloth : alk. paper)
1. Water quality—Management—Statistical methods. 2. Water quality management—Statistical methods. I. Title.
TD367.M34 2005
628.1'61—dc22 2004059090

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

*This book is dedicated to*

*Robyn*
*Stephen and Christine*
*Richard, Chris*
*Desmond*

This Page Intentionally Left Blank

# *Contents*

**List of Figures** xv

**List of Tables** xix

**Preface** xxi

*Part I Issues*

**1 Introduction** **1**

*1.1 Conventions* *1*

*1.2 The essentials* *2*

*1.3 Meeting management's information needs* *4*

*1.4 Water quality observations as random variables* *5*

*1.5 Samples and populations* *6*

*1.6 Special characteristics of water quality data* *7*

*1.7 Data analysis protocols* *9*

**2 Basic concepts of probability and statistics** **11**

2.1 Probability rules 13
2.2 Representing data 16
2.2.1 Types of data 16
2.2.2 Frequency, bar graph, and histogram 17
2.2.3 Describing the distribution of probability 18
2.2.4 Discrete versus continuous data 19
2.2.5 Summary statistics 20
2.3 Exploratory and graphical methods 22
2.4 Important distributions 25
2.5 Continuous distributions 27
2.5.1 Normal distribution 27
2.5.2 Lognormal distribution 30
2.5.3 Gamma distribution 31
2.5.4 Beta distribution 32
2.6 Discrete distributions 33
2.6.1 Binomial distribution 33
2.6.2 Poisson distribution 36
2.6.3 Negative binomial distribution 36
2.6.4 Hypergeometric distribution 36
2.6.5 Multinomial distribution 36
2.7 Sampling distributions 36
2.7.1 Student's t-distribution 37
2.7.2 Chi-square distribution 37
2.7.3 F-distribution 38
2.8 Statistical tables 38
2.9 Correlation and measures thereof 41
2.10 Statistical models and model parameters 42
2.11 Serial correlation, seasonality, trend, and scale 43
2.11.1 Effect of serial correlation 44
2.12 Regression 47
2.12.1 Applications to water quality 48
2.12.2 Nonparametric regression 49
2.13 Estimating model parameters 49
2.13.1 Point versus interval estimation 49

2.13.2 *Interval estimates* 50
2.13.3 *Bias* 51
2.13.4 *Percentiles* 51
*Problems* 52
*Appendix: Conditional probabilities—The Monty Hall dilemma* 54

**3 Intervals** **57**
3.1 *Confidence intervals* 59
3.1.1 *For means* 59
3.1.2 *For prediction* 66
3.1.3 *For percentiles* 68
3.2 *Tolerance intervals* 70
3.3 *Credible intervals* 71
3.4 *Intervals literature* 72
*Problems* 72

**4 Hypothesis testing** **75**
4.1 *The classical approach* 76
4.2 *The main steps in the classical approach* 76
4.2.1 *What is meant by a decision?* 78
4.3 *Commentary on the five steps* 79
4.3.1 *Step 1: State the hypothesis to be tested* 79
4.3.2 *Step 2: State the acceptable error risk* 82
4.3.3 *Step 3: Sampling* 83
4.3.4 *Step 4: Apply the decision rule* 83
4.3.5 *Step 5: Rejecting, and accepting (maybe!)* 86
4.4 *An "evidence-based" approach (for dichotomous data)* 93
4.5 *Power analysis* 94
4.6 *Multiple comparisons* 95
4.7 *Nonparametric tests* 98
4.8 *Randomization tests* 100
4.9 *Bayesian methods* 101

4.10 *Likelihood approach* 102
4.11 *Hypothesis testing literature* 103
4.11.1 *The point-null hypothesis testing debate* 103
4.11.2 *The p-value culture* 104
*Problems* 105

**5 Detection** **107**
5.1 *One-sided hypothesis tests* 108
5.1.1 *Defining detection probability* 108
5.1.2 *What burden of proof?* 110
5.1.3 *Decision rules and detection probabilities* 111
5.1.4 *Detection regions* 113
5.1.5 *Non-normality and unequal variances* 114
5.2 *Point-null hypothesis tests* 115
5.2.1 *Defining detection probability* 115
5.2.2 *Power depends on sample size* 116
5.2.3 *Decision rules and detection probabilities* 118
5.2.4 *Detection regions* 119
5.3 *Equivalence tests* 120
5.3.1 *Defining detection probability* 120
5.3.2 *Decision rules and detection probabilities* 123
5.3.3 *Equivalence tests versus point-null tests* 125
5.3.4 *Detection regions* 126
5.3.5 *Equivalence versus the Power Approach* 128
5.3.6 *A potted history of equivalence testing* 130
5.4 *p-value behavior under different hypotheses* 131
5.5 *What has been detected?* 132
5.6 *Tests and confidence intervals* 132
5.6.1 *A Bayesian alternative* 134
5.7 *Association or causation?* 134
*Problems* 136

**6 Mathematics and calculation methods** **139**
6.1 *Introduction* 139

6.2 *Formulae and calculation methods for important distributions* 140
- 6.2.1 *Normal distribution* 140
- 6.2.2 *Lognormal distribution* 141
- 6.2.3 *Gamma distribution* 141
- 6.2.4 *Beta distribution* 142
- 6.2.5 *Binomial distribution* 143
- 6.2.6 *Poisson distribution* 144
- 6.2.7 *Negative binomial distribution* 145
- 6.2.8 *Hypergeometric distribution* 147
- 6.2.9 *Multinomial distribution* 148
- 6.2.10 *Student's t-distribution* 148
- 6.2.11 *Chi-square distribution* 149
- 6.2.12 *F-distribution* 149
- *Problems* 150

6.3 *Calculating confidence and tolerance limits on normal percentiles* 152
- 6.3.1 *Two-sided equi-tailed confidence limits on normal percentiles* 152
- 6.3.2 *One-sided confidence (tolerance) limits on normal percentiles* 154
- 6.3.3 *Two-sided tolerance limits on percentiles* 154
- 6.3.4 *Calculating noncentral t-probabilities* 155
- *Problems* 156

6.4 *Development of the detection formulae* 157
- 6.4.1 *One-sided test, unknown variance* 157
- 6.4.2 *One-sided test, known variance* 158
- 6.4.3 *Point-null hypothesis test, unknown variance* 158
- 6.4.4 *Point-null hypothesis test, known variance* 159
- 6.4.5 *Equivalence hypothesis, known variance* 159
- 6.4.6 *Inequivalence hypothesis, known variance* 161
- 6.4.7 *(In)equivalence hypotheses, unknown variance* 162

6.4.8 *Detection regions for equivalence tests and the Power Approach* 163
6.4.9 *Calculating the incomplete noncentral t-probabilities* 165
*Problems* 167

*Part II Problems and Solutions*

**7 Formulating water quality standards** **171**
7.1 *Setting the scene* 171
7.1.1 *Explicit and narrative statements* 172
7.2 *Key questions* 173
7.3 *Enforceability* 174
7.4 *Requirements for enforceability* 175
*Problems* 176

**8 Percentile standards (and the Reverend Bayes)** **179**
8.1 *Two forms of percentile standards* 179
8.2 *Calculating nonparametric sample percentiles* 180
8.2.1 *Comparing the alternatives* 181
8.2.2 *Median versus geometric mean* 182
8.3 *Percentiles versus maxima* 182
8.4 *Percentile compliance rules* 183
8.4.1 *Developing rules using the Classical approach* 184
8.4.2 *Developing rules using the Bayesian approach* 185
8.4.3 *Comparing the Classical and Bayesian results* 190
8.4.4 *Conclusion* 192
*Problems* 193

**9 Microbial water quality and human health** **195**
9.1 *The key elements* 196
9.1.1 *Hazard identification* 196
9.1.2 *Exposure assessment* 198
9.1.3 *Dose-response* 201

9.1.4 *Risk profiling* 203

9.2 *Risk communication* 204

9.3 *Some useful mathematical details* 206

9.3.1 *Added-zeroes distributions for dichotomous data* 206

9.3.2 *Hockey stick distribution* 207

9.3.3 *Loglogistic distribution* 208

9.3.4 *Dose-response curves* 208

9.3.5 *Conditional dose-response relationships* 210

9.3.6 *What if the distribution of pathogens is not Poisson?* 211

9.3.7 *Comparing conditional and average-dose risks* 212

*Problems* 213

**10 MPNs and microbiology** **215**

10.1 *MPNs in standard tables* 216

10.2 *MPN "confidence intervals"* 218

10.3 *Sampling an MPN distribution* 218

10.4 *Calculating exact MPNs* 222

10.4.1 *Considering each set separately* 222

10.4.2 *Combining the sets* 223

*Problems* 224

**11 Trends, impacts, concordance, detection limits** **225**

11.1 *Detecting and analyzing trends* 225

11.1.1 *Trend slope estimator* 225

11.1.2 *Seasonal Kendall trend test* 226

11.1.3 *Two anomalies* 227

11.1.4 *Multi-site trends; a Bayesian approach* 228

11.2 *Impacts and nominal data* 231

11.3 *Concordance assessment* 232

11.3.1 *Cohen's kappa for dichotomous variables* 232

11.3.2 *Continuous variables—Lin's concordance correlation coefficient* 236

11.4 *Detection limits* 236
11.4.1 *Type A censoring* 236
11.4.2 *Type B censoring* 237
11.4.3 *Handling "less than" data* 238
*Problems* 239
*Appendix* 240

**12 Answers to exercises** **243**

**References** **271**

**Author Index** **295**

**Topic Index** **299**

**Appendix A Statistical Tables** **305**
A.1 *Normal distribution* 306
A.1.1 *Cumulative unit normal distribution* 306
A.1.2 *Percentiles of the unit normal distribution* 307
A.2 *t-distribution* 308
A.2.1 *Cumulative Student's t-distribution* 308
A.2.2 *Percentiles of Student's t-distribution* 309
A.3 *Binomial distribution* 310
A.3.1 *Cumulative binomial probabilities* 310
A.4 *Normal confidence and tolerance intervals* 311
A.4.1 *k factor for two-sided 95% confidence intervals on a normal percentile* 311
A.4.2 *k factor for one-sided confidence (or tolerance) intervals on a normal percentile* 312
A.4.3 *k factor for two-sided tolerance intervals on a normal percentile* 313

# *List of Figures*


| | | |
|---|---|---|
| *1.1* | *Accurate observations are both precise and unbiased.* | *7* |
| *2.1* | *Bar graph of MPN data.* | *18* |
| *2.2* | *Cumulative distribution function of MPN data.* | *19* |
| *2.3* | *Symmetry and skewness.* | *23* |
| *2.4* | *Lowess fit through data for the Ngakaroa Stream.* | *24* |
| *2.5* | *Boxplot for somatic coliphage in recreational waters.* | *25* |
| *2.6* | *Probability density function (pdf) for the unit normal distribution.* | *28* |
| *2.7* | *Cumulative distribution function (CDF) for the unit normal distribution.* | *29* |
| *2.8* | *Normal and lognormal probability density functions.* | *31* |
| *2.9* | *A variety of shapes for the pdf of the two-parameter gamma distribution.* | *32* |
| *2.10* | *A variety of shapes for the pdf of the two-parameter beta distribution.* | *33* |

2.11 *Probability mass functions for three common discrete distributions.* 35

2.12 *Student t-distributions and the unit normal distribution.* 37

2.13 *Probability density functions for the* $\chi^2$*- and F-distributions.* 38

2.14 *Five possibilities for reporting areas under the t-distribution.* 39

2.15 *Possible linear correlations.* 41

2.16 *Field Raynes effluent suspended solids data and serial correlation structure.* 45

3.1 *Geometric mean confidence limits as a function of sample size.* 62

3.2 *"Error bars."* 66

3.3 *Prediction intervals and confidence intervals for linear regression.* 68

3.4 *Confidence limits for the 95%ile as a function of sample size.* 69

3.5 *Highest density and equi-tails regions.* 72

4.1 *Critical region for a one-sided test using the unit normal distribution.* 84

4.2 *Critical region for a two-sided test using the unit normal distribution.* 86

4.3 *Binomial distribution for* $n = 20$ *and* $\theta = 0.5$. 89

4.4 *Negative binomial distribution for* $k = 5$ *and* $\theta = 0.5$. 90

4.5 *Binomial distribution for* $n = 40$ *and* $\theta = 0.5$. 92

4.6 *ANOVA: inferring differences in means from differences in variances.* 96

5.1 *Detection curves for two sample sizes for a one-sided inferiority test.* 110

5.2 *Probabilities of accepting and rejecting complementary one-sided hypotheses.* 111

5.3 *Detection regions for inferiority and superiority tests.* 115

5.4 *Detection curve for the point-null hypothesis test.* 116

5.5 *Effect of increasing sample size on detection curves for a point-null test.* 117

5.6 *Detection region for the point-null hypothesis.* 119

5.7 *Detection probability for testing the equivalence hypothesis.* 121

5.8 *Detection probabilities for testing the equivalence and inequivalence hypotheses (variance unknown).* 122

5.9 *Detection probabilities for testing the equivalence and inequivalence hypotheses (variance known).* 123

5.10 *"Two one-sided test" definitions in terms of confidence intervals.* 124

5.11 *Detection curves for equivalence vs. point-null tests.* 127

5.12 *Detection regions for testing the inequivalence and equivalence hypotheses.* 128

5.13 *Detection regions for testing the inequivalence and equivalence hypotheses, compared to the Power Approach.* 129

6.1 *Unit normal, Student's t, and noncentral t pdfs.* 153

8.1 *Ratio of 95%ile to maximum for E. coli at 317 sites.* 183

8.2 *Maximum and 95%ile E. coli at 317 sites.* 183

8.3 *Four prior densities compared.* 188

8.4 *"Confidence of Compliance" with a 95%ile: sensitivity to prior distributions.* 190

8.5 *Confidence of Compliance for five percentiles.* 191

9.1 *General QRA flow diagram.* 197

9.2 *Fitting shellfish meal size data.* 200

9.3 *Virus and bacteria dose-response models.* 202

9.4 *"Hockey stick" empirical influent virus distribution.* 207

9.5 *Comparing simple exponential and binomial models.* 212

9.6 *Comparing beta-Poisson and beta-binomial models.* 213

10.1 *Binning procedure for MPN data.* 221

11.1 *Lowess fit through flow-adjusted data for the Ngakaroa Stream.* 229

11.2 *Robust fill-in for "less than" data.* 239

# *List of Tables*


2.1 *Most probable number (MPN) data for shellfish-growing waters* 17

2.2 *Summary of terms for sample data* 21

2.3 *Summary of terms for populations* 21

4.1 *Matrix of possible results from a hypothesis test* 77

5.1 *Definition of population variables* 112

5.2 *Inferiority and superiority hypotheses* 112

5.3 *Decision rules and detection probabilities for inferiority and superiority tests* 113

5.4 *Formulae for inferiority and superiority tests and for point-null tests* 114

5.5 *Sample size and confidence interval width* 117

5.6 *Hypotheses for point-null hypothesis tests* 118

5.7 *Decision rules and detection probabilities for point-null hypothesis tests* 118

| | | |
|---|---|---|
| 5.8 | *Equivalence and inequivalence hypotheses* | 124 |
| 5.9 | *Decision rules and detection probabilities for equivalence hypothesis tests* | 125 |
| 5.10 | *Formulae for equivalence tests* | 126 |
| 5.11 | *Criteria for evidence of association* | 136 |
| 8.1 | *Permissible exceedances in a proof of safety approach for a 95%ile standard* | 192 |
| 8.2 | *Permissible exceedances in a proof of hazard approach for a 95%ile standard* | 193 |
| 9.1 | *Coastal shellfish bioaccumulation factors for two seasons (USA)* | 200 |
| 9.2 | *Calculated Campylobacter infection rates* | 204 |
| 10.1 | *Comparing MPN values* | 217 |
| 10.2 | *Comparing MPN confidence and credible limits.* | 219 |
| 10.3 | *MPN occurrence probabilities.* | 220 |
| 11.1 | *Groundwater trends in Canterbury* | 229 |
| 11.2 | *Presence/absence table* | 233 |
| 11.3 | *Frequency table for presence/absence data* | 233 |
| 11.4 | *Chance-correction table for presence/absence data* | 233 |
| 11.5 | *Suggested agreement descriptors.* | 234 |

# *Preface*

When managing water resources we often seek to separate pattern from randomness. For example, is the quality of water in a lake really deteriorating? Is that because the land-use patterns in its watershed are intensifying? Just how sure can we be of a conclusion about that? What are the health risks from swimming in this river or at that beach? Is there an important change in stream benthic communities downstream of mine waste discharges? These are issues that statistical methods can shed light on. I hope this book will help water resources managers and scientists to formulate, implement and interpret better and more appropriate methods to address such matters.

To do so, I have not sought to repeat material readily available in other texts.[1] Rather than being a "how to" compendium of the many procedures to be found in them, I have focused on material not generally available elsewhere, both in general (such as in Bayesian modes of inference) and in particular (such as percentile standards, water-related human health risk modeling, and MPN methodology for microbiological enumerations).

This book's origins lie in an invitation from my colleague, Dr Bryan Manly, to present a paper to the Joint Statistical Meetings held in 2000 at Indianapolis. The title of the paper (as suggested by him) was "Statistical Methods Helping and Hindering Environmental Science and Management." It drew on experience I had gained over some years in using statistical methods to address practical questions that commonly

[1]Green (1979 [127]), Berthouex & Brown (1980 [23]), Snedecor & Cochran (1980 [291]), Sokal & Rohlf (1981 [292]), Iman & Conover (1983 [161]), Gilbert (1987 [118]), Krebs (1989 [176]), Ward *et al.* (1980 [338]), Gibbons (1994 [117]), Zar (1996 [349]), Manly (2001 [198]), Millard & Neerchal (2001 [226]), Helsel & Hirsch (2002 [147]), Townend (2002 [319]), Bolstad (2004 [28]).

arise in managing water resources. For example: Just what does the analyst mean when stating that a result is "statistically significant"? How should one devise and document percentile standards for effluent discharges, drinking water, or environmental standards? How many samples should be required in drinking water standards in order to provide a satisfactory level of assurance that the waters' quality sufficiently protects public health? When and how can one take account of what we think we already know when framing compliance rules or when analyzing experiments?

The Indianapolis paper, with its somewhat provocative title, was subsequently published (McBride 2002 [206]), and it resulted in an invitation to write this book. In doing so, I have attempted to give "lines of approach" (rather than "answers") to a range of issues that arise in water management (and in water science), such as those given above. I say "lines of approach" very deliberately because I have so often been asked, "What is the statistically correct way of analyzing these data?" In fact, there is seldom (if ever) *one* correct way, either for analysis of data or for design of sampling programs. We make judgments, not rules (Stewart-Oaten 1995 [300]). There may be a number of alternative approaches, each with its merits and drawbacks, and awareness of these can only be beneficial. There are of course many more inappropriate and incorrect lines of approach. So a function of this book is to clarify the appropriateness of such methods for various issues. By doing so, I hope to alert scientists and water resources managers to the wide variety of fruitful statistical methods that can be used. For example, it can come as a surprise that statistical methods do not just deal with numbers; incorporation of nominal data can greatly enhance their utility (e.g., using water color, wind direction, and octants of cloud cover in models of near-shore contamination by fecal bacteria).

In planning this book, I had first thought that it would be wise to introduce statistical concepts gradually, by way of a series of practical examples. As an instance of that, one could introduce Bayesian approaches to data analysis when considering the framing of percentile compliance rules. However, on reflection and discussion with colleagues, it became apparent that some general and pervasive issues needed first to be discussed in their own right. Accordingly, the book is in two parts. The first, *Issues*, consisting of six chapters, presents the material I consider to be important to be understood when contemplating using statistical methods for water resources management. It includes a number of practical examples. The five chapters in the second part, *Problems and Solutions*, draw on material given in the first part. They cover a range of topics that often arise but are not covered in most texts (formulating environmental standards, using percentile standards, microbiological water quality, "most probable numbers" (MPNs) and human health, and a catch-all chapter including material on trends, impacts, concordance and detection limits). Each chapter contains a number of set problems, for which full "answers" are give in the final (12th) chapter. Some of these problems are more difficult than others; some are discursive, in which case "answers" are longer than is usual in a statistics text.

So what are those "general and pervasive issues"? Some are details—but important details (especially concerning correct usage of standard errors and explaining "error bars"). Others are substantial. They have to do with identifying management's information needs, the role of various forms of hypothesis testing, and the types of intervals that may be used to account for uncertainty (confidence intervals, tolerance intervals, and credible intervals). To set the scene for these issues, Chapter 1 discusses the use of statistical populations and samples in water management. Chapter 2 presents basic concepts of probability and statistics, including different modes of statistical inference, especially frequentist versus Bayesian (throughout the book I present the case for using both). Chapter 3 discusses and contrasts confidence intervals, tolerance intervals and credible intervals. Chapters 4 and 5 then cover general and specific issues to do with statistical hypothesis testing—these tests are widely used in water management, and not always appropriately. These chapters address questions such as: When are "one-sided" tests appropriate? What can we actually infer from a test of a single "null" hypothesis? When should we use nonparametric approaches? Is there a corresponding Bayesian test? What is the role of equivalence testing? How do we give effect to the precautionary approach? We also introduce, in Chapter 5, the "detection probability" in the context of one-sided and two-sided tests. Detailed (though not rigorous) mathematical arguments and calculation routines are presented in Chapter 6—finer mathematical details are admirably covered in papers and other texts (e.g., Ferguson 1967 [90], Lehmann 1986 [183], Freund 1992 [101], Lee 1997 [179], Gelman *et al.* 2000 [110], Casella & Berger 2002 [45], Wellek 2003 [340]). This chapter can be skipped without losing too much of the main information presented.

A number of these issues are seldom addressed in much detail in applied statistical texts. For example, most workers actually interpret confidence intervals in a Bayesian manner, yet the Bayesian view of probability and associated modes of inference are seldom presented. It is important for environmental professionals to grasp such matters and their consequences. I became aware of this in 1989 when, while browsing in the basement of the Colorado State University library, I stumbled across the 1970 text by Morrison & Henkel *The Significance Test Controversy* ([232]). My response was one of surprise: "What controversy?" I read avidly, from a text arising not from environmental science, but from psychology (from where so many statistical advances have emanated). Many colleagues involved in water science and management (in my home country and abroad) have had the same response when confronted by the notion that statistical methodology, especially hypothesis testing, is accompanied (to this day) in the statistical literature by a degree of controversy. Some statisticians have even advocated that tests be abandoned altogether [e.g., in chapters of a recent book, edited by Harlow, Muliak, and Steiger (1997 [136]): *What if there were no significance tests?*]. But, in water management at least, we cannot abandon tests—comparisons, hence tests, are often required (between sites, between treatment levels, with standards,. . .).

Throughout I have made use of footnotes, giving detail that readers can pass over. More substantial technical details are contained in Appendices to some chapters. These too can be passed over without compromising understanding of the text. References are contained in a single list at the book's end, along with an author index and a subject index.

The book's text was prepared using LATEX, using the versatile WinEdt software. Most of the graphics have been prepared using Kaleidagraph$^{TM}$. A lot of the text was prepared listening to Miles Davis' *In a Silent Way.*

*Acknowledgements* Firstly, I thank Julian Ellis (Water Research Centre, England), whose "NS 29" deserves to be more widely known and read (see reference [84]), Robert Ward and Jim Loftis at Colorado State University, and Bryan Manly at West Inc., Wyoming (formerly at Otago University, New Zealand). Some of the material herein has been derived from Short Courses that Jim, Robert, Bryan and I have taught together. Other Fort Collins colleagues (Tom Sanders at CSU and Roy Irwin at the US National Park Service) have been a stimulation, as have many colleagues at NIWA (the National Institute of Water and Atmospheric Research)—Russell Cole, Bryce Cooper, Rob Collins, Rob Davies-Colley, Chris Francis, Mal Green, Judi Hewitt, Ian Jowett, Charles Pearson, Rick Pridmore, Kit Rutherford, Mike Scarsbrook, Rebecca Stott, James Sukias, Simon Thrush and Bob Wilcock. Another NIWA colleague, Mark Meleason, made a comprehensive and enthusiastic review of draft chapters as did also Alec Zwart (University of Waikato, Department of Statistics, Hamilton, New Zealand) and David Smith (Bureau of Water Supply, New York City Department of Environmental Protection). Murray Jorgensen, Bill Bolstad, James Curran, and Judith McWhirter (University of Waikato) have provided insights and feedback, as have Michael Taylor and Alexander Kouzminov (Ministry of Health, New Zealand). I also benefitted from discussions of many issues with other colleagues, both in New Zealand (Stan Abbott, Trevor Atkins, Andrew Ball, Noel Burns, James Court, Chris Hatton, Rob Lake, Gillian Lewis, Megan Linwood, Dorothy McCoubrey, John Maxted, Chris Nokes, Terry Ryan, Chris Skelly, Brett Stansfield, Desmond Till, Cliff Tipler, Peter van der Logt and Bill Vant), and abroad (Steve Hrudey, Dennis Helsel, Chuck Haas, Joe Germano, Sander Greenland, David Parkhurst and Samantha Rizak).

My employer (NIWA) has been very supportive. Funding for some of the work presented here has come from the New Zealand Foundation for Research, Science and Technology, and also from the New Zealand Ministry for the Environment, the Ministry of Health, and the Ministry of Research, Science and Technology. John Moore (Christchurch City Council) made coastal outfall human health risk analysis results available (for Chapter 9). Chris Hatton and Grant Barnes (Auckland Regional Council) made the Ngakaroa Stream data available (for Chapters 2 and 11). Bruce Apperley (Gisborne District Council, New Zealand) and Carl Hanson (Environment Canterbury, Christchurch, New Zealand) provided data cited in Chapter 11. Joke Baars (NIWA's librarian) has been unfailing in her industry, getting many obscure books and articles. Alison Bartley (NIWA) typed endless citations. I thank Susanne Steitz, Lisa Van Horn, Heather Bergman and Stephen Quigley at Wiley, and especially Amy Hendrickson at TeXnology Inc., for their unfailing help in producing the text.

Finally, I thank my wife Robyn for her constant support and encouragement.

GRAHAM B. MCBRIDE

*Hamilton, New Zealand*

# *Part I*

# *Issues*

'Far better an approximate answer to the right question, which is often vague, than the exact answer to the wrong question, which can always be made precise.'

—John Tukey (1962 [321])

'It's all in the mind, you know.'

—Wallace Greenslade and Spike Milligan, *The Goon Show* (1952–1960)

This Page Intentionally Left Blank

# 1

# *Introduction*

## 1.1 CONVENTIONS

Key statistical terms are written in **bold type** when they first appear. Italicized upper case Roman letters are used to denote data and the statistics derived from them, making it easier to distinguish them from text. For example $X_i$ refers to the *i*th numerical **datum** in a **sample**, and $\bar{X}$ is their **arithmetic mean**—usually just stated as the **mean**.[1] Data do not have to be numbers, they can be **categories** (such as water color), but in that case a mean is not defined. Italicized lowercase Greek letters are used for **population parameters** such as the **true mean**, which is **estimated** by $\bar{X}$ (so the true mean is denoted by $\mu$). Samples are **drawn** from populations, and the parameters are a succinct way to summarize the shape of the populations. Lowercase italicized Roman letters (such as $x$ and $y$) refer to the population **random variables** from which samples may be drawn. Particular values of those random variables are written in uppercase letters (e.g., $X$ and $Y$). $X_p$ represents the $100p$th percentile of the $x$ distribution (so $X_{0.95}$ is the 95th percentile—hereafter written as "95%ile"). Overbars always denote a mean; for example, the difference between the means of two **independent** samples drawn from $x$ and $y$ is written as $\bar{X} - \bar{Y}$, whereas if those samples were **paired** we would be contemplating the mean difference $\overline{X - Y}$.

Like any discipline, statistics has its own language conventions. While most can be explained as we go along, those to do with the words **sample** and **error**, the symbol ***p***, and even the word **statistics** need some explanation (though the context should make the meaning of such phrases clear).

[1]Other types of means may be used, especially the **geometric mean**.

To an environmental professional a sample is a volume or mass of material taken from the environment—for example, a container of stream water for subsequent phosphorus analysis in a laboratory, or a reading of pH from a probe. But to a statistician, a sample is a collection of results or **observations**—for example, phosphorus concentrations in a set of water "samples." So to a statistician one **sample** contains **many data** and the **sample size** is the number of data in the sample—not the volume of the container. Similarly, an environmental professional would regard an **error** as just that—a mistake. But to a statistician, **sampling error** is the natural **variability** inherent among data taken from a **population** and is therefore always present and to be accounted for. If there are errors in measurement (nothing in this world is perfect), these are referred to as **measurement error**.

Then we have a **standard error**. This is not a recommended way to make a mistake! It is another measure of variability—not of the data, but of a parameter estimated from the data, such as the mean.[2] Next, because of conventions in wide use, we are forced to use the symbol $p$ ambiguously (it can mean **$p$-value** (an **exceedance probability**) or **proportion**. Uppercase $P$ generally denotes a **cumulative probability**, expressed as a percentage (e.g., the $P$th percentile).

Finally, the term "**statistics**" itself can have two meanings. The singular **statistic** is a number that characterizes some feature of a set of data—for example, the sample arithmetic mean as a measure of the **central tendency** of data. The plural **statistics** can also refer to the methods that use such numbers to make estimates and comparisons. In this text we generally use the term **statistic** for the former meaning and **statistical methods** for the latter.

## 1.2 THE ESSENTIALS

The measurements we make are taken from samples that are only a tiny proportion of the water body that we want information about. So any **inferences** drawn from the data are **uncertain** to some degree. Statistical methods are a means of handling this uncertainty, both in designing sampling programs and in obtaining useful information from the results of sampling.

**Classical (frequentist)** statistical methods allow us to draw inferences from sample data with, under certain assumptions, a known degree of uncertainty. Furthermore, they can permit different data analysts to follow the same procedures, thus reaching the same conclusions from a given set of data.[3] Statistical methods also provide a means of specifying how monitoring data should be analyzed before they are collected, thus ensuring that the information needs of management will be met from a planned monitoring program.

[2]If that parameter is the mean the standard error is in fact the **standard deviation** *of the mean* of the data, *not* of the data themselves. Importantly, it tends to get smaller as the number of data increases. More on that later (Section 2.13.1).

[3]This is not the case for the **Bayesian** methods that we shall meet later—although these can have some definite advantages, as we shall see.