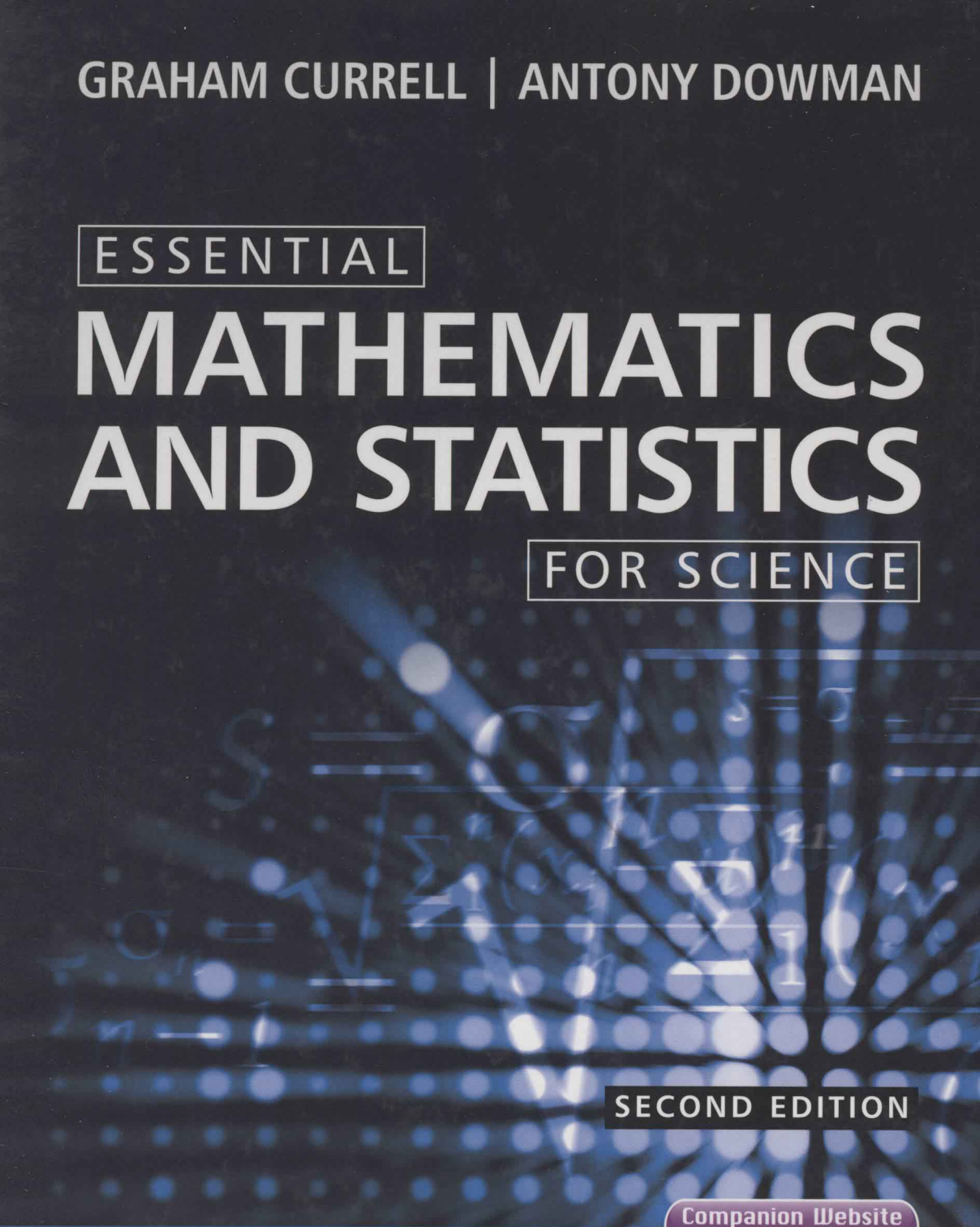
GRAHAM CURRELL | ANTONY DOWMAN

# ESSENTIAL MATHEMATICS AND STATISTICS FOR SCIENCE



SECOND EDITION

Companion Website
Multi-media Study

WILEY-BLACKWELL

# Essential Mathematics and Statistics for Science Second Edition

**Graham Currell**

**Antony Dowman**

*The University of the West of England, UK*

WILEY-BLACKWELL

A John Wiley & Sons, Ltd., Publication

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical and Medical business with Blackwell Publishing.

*Registered office:* John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

*Other Editorial offices:*
9600 Garsington Road, Oxford, OX4 2DQ, UK

111 River Street, Hoboken, NJ 07030-5774, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.



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

Currell, Graham.
Essential mathematics and statistics for science / Graham Currell, Antony Dowman. – 2nd ed.
p. cm.
Includes index.
ISBN 978-0-470-69449-7 – ISBN 978-0-470-69448-0
1. Science–Statistical methods. 2. Science–Mathematics. I. Dowman, Antony. II. Title.
Q180.55.S7C87 2009
507.2–dc22

2008052795

ISBN: 978-0-470-69449-7 (HB)
978-0-470-69448-0 (PB)

A catalogue record for this book is available from the British Library.

Typeset in 10/12pt Times and Century Gothic by Laserwords Private Limited, Chennai, India.
Printed and bound in Great Britain by CPI Antony Rowe, Chippenham, Wiltshire.

First Impression 2009

# Essential Mathematics and Statistics for Science Second Edition

*To*

*Jenny and Felix*

*Jan, Ben and Jo.*

# Preface

The main changes in the second edition have been driven by the authors' direct experience of using the book as a core text for teaching mathematics and statistics to students on a range of undergraduate science courses.

Major developments include:

- Integration of 'how to do it' *video clips* via the Website to provide students with audio-visual worked answers to over 200 'Q' questions in the book.
- Improvement in the *educational development* for certain topics, providing a greater clarity in the learning process for students, e.g. in the approach to handling equations in Chapter 3 and the development of exponential growth in Chapter 5.
- Reorientation in the approach to hypothesis testing to give priority to an understanding of the *interpretation of p-values*, although still retaining the calculation of test statistics. The statistics content has been substantially reorganized.
- Movement of some *content to the Website*, e.g. Bayesian statistics and some of the statistical theory underpinning regression and analysis of variance.
- Revised *computing tutorials* on the Website to demonstrate the use of Excel and Minitab for many of the data analysis techniques. These include *video* demonstrations of the required keystrokes for important techniques.

The book was designed principally as a study text for students on a range of undergraduate science programmes: biological, environmental, chemical, forensic and sports sciences. It covers the majority of mathematical and statistical topics introduced in the first two years of such programmes, but also provides important aspects of experimental design and data analysis that students require when carrying out extended project work in the later years of their degree programmes.

The comprehensive Website actively supports the content of the book, now including extensive video support. The book can be used independently of the Website, but the close integration between them provides a greater range and depth of study possibilities. The Website can be accessed at:

www.wiley.com/go/currellmaths2

The introductory level of the book assumes that readers will have studied mathematics with moderate success to Year 11 of normal schooling. Currently in the UK, this is equivalent to a Grade C in Mathematics in the General Certificate of Secondary Education (GCSE).

There are Revision Mathematics notes available on the associated Website for those readers who need to refresh their memories on relevant topics of basic mathematics – BODMAS, number line, fractions, percentages, areas and volumes, etc. A self-assessment test on these

'basic' topics is also available on the Website to allow readers to assess their need to use this material.

The first eight chapters in the book introduce the *basic* mathematics and statistics that are required for the modelling of many different scientific systems. The remaining chapters are then primarily related to *experimental investigation* in science, and introduce the statistical techniques that underpin data analysis and hypothesis testing.

Over 200 worked Examples in the text are used to develop the various topics. The calculations for many of these Examples are also performed using Microsoft Excel (office.microsoft.com) and the statistical analysis program Minitab (www.minitab.com). The files for these calculations are available via the Website.

Readers can test their understanding as each topic develops by working through over 200 'Q' questions in the book. The numeric answers are given at the end of the book, but full *worked* answers are also available through the Website in both video and printed (pdf) format.

Throughout the book, readers have the opportunity of learning how to use software to perform many of the calculations. This strong integration of paper-based and computer-based calculations both supports an understanding of the mathematics and statistics involved and develops experience with the use of appropriate software for data handling and analysis.

## Scientific context

The diverse uses of mathematics and statistics in the various disciplines of science place different emphases on the various topics. However, there is a core of mathematical and statistical techniques that is essentially common to all branches of experimental science, and it is this material that forms the basis of this book. We believe that we have developed a coherent approach and consistent nomenclature, which will make the material appropriate across the various disciplines.

When developing questions and examples at an introductory level, it is important to achieve a balance between treating each topic as pure mathematics or embedding it deeply in a scientific 'context'. Too little 'context' can reduce the scientific interest, but too much can confuse the understanding of the mathematics. The optimum balance varies with topic and level.

The 'Q' questions and Examples in the book concentrate on clarity in developing the topics step by step through each chapter. Where possible we have included a scientific context that is understandable to readers from a range of different disciplines.

## Experimental design

The process of good experimental planning and design is a topic that is often much neglected in an undergraduate course. Although the topic pervades all aspects of science, it does not have a clear focus in any one particular branch of the science, and is rarely treated coherently in its own right.

Good experimental design is dependent on the availability of suitable mathematical and statistical techniques to analyse the resulting data. A wide range of such methods are introduced in this book:

- Regression analysis (Chapters 4 and 13) for relationships that are inherently linear or can be linearized.

- Logarithmic and/or exponential functions (Chapter 5) for systems involving natural growth and decay, or for systems with a logarithmic response.
- Modelling with Excel (Chapter 6) for rates of change.
- Probabilities (Chapter 7), frequency and proportions (Chapter 14) and Bayesian statistics (Website) to interpret categorical data, ratios and likelihood.
- Statistical distributions (Chapter 8) for modelling random behaviour in complex systems.
- Statistical analysis (Chapters 9 to 14) for hypothesis testing in a variety of systems.
- Analysis of variance (Chapter 11) for hypothesis testing of complex experimental systems.
- Experimental design overview (Chapter 15).

# Computing software

There are various software packages available that can help scientists in implementing mathematics and statistics. Some university departments have strong preferences for one or the other.

Microsoft Excel spreadsheets can be used effectively for a variety of purposes:

- basic data handling – sorting and manipulating data;
- data presentation using graphs, charts, tables;
- preparing data and graphs for export to other packages;
- performing a range of mathematical calculations; and
- performing a range of statistical calculations.

Minitab (Minitab Inc.) is designed specifically for statistical data analysis. The data is entered in columns and a wide range of analyses can be performed using menu-driven instructions and interactive dialogue boxes. The results are provided as printed text, graphs or new column data.

Most students find that the statistical functions in Excel are a helpful *introduction* to using statistics, but for particular problems it is more useful to turn to the packages designed specifically for statistical analysis. Nevertheless, it is usually convenient to use Excel for organizing data into an appropriate layout before exporting to the specialized package.

The book has used Excel 2003 and Minitab 15 to provide all of the software calculations used, and the relevant files are available on the Website. However, there are several other software packages that can perform similar tasks, and information on some of these is also given on the Website.

Most of the graphs in the book have been prepared using Excel, except for those identified as having been produced using Minitab.

# On-line Learning Support

The book's Website (www.wiley.com/go/currellmaths2) provides extensive learning support integrated closely with the content of the book.

Important learning elements referenced *within* the book are:

- **Examples** (e.g. **Example 7.12**) with worked answers given directly within the text, and with supporting files available on the Website where appropriate.
- **'Q' questions** (e.g. **Q7.13**) with *numerical* answers at the end of the book, but with *full worked* answers on video or pdf files via the Website.
- **Equations** – referred to using **square brackets**, e.g. **[7.16]**.

The Website for the second edition provides the following structural support:

- **'How to do it' – answers to all 'Q' questions**. Over 200 flash video clips provide worked answers to all of the 'Q' questions in the book, and can be viewed directly over the Internet. The worked answers are also presented in pdf files.
- **Further practice questions**. Additional questions and answers are provided which enable students to further practise/test their understanding. Many students find these particularly useful in some *skill* areas, such as chemical calculations, rearranging equations, logs and exponentials, etc.
- **Excel and Minitab tutorials**. Keystroke tutorials provide a guide to using Excel 2007 and Minitab 15 for some of the important analyses developed in the book.
- **Excel and Minitab files**. These files provide the software calculations for the examples, 'Q' questions, tables and figures presented in the book. In appropriate cases, these are linked with video explanations.
- **Additional materials**. Additional learning materials (pdf files), including revision mathematics (basic skills of the number line, BODMAS, fractions, powers, areas and volumes), Bayesian statistics, transformation of data, weighted and nonlinear regression, data variance.
- **Reference materials**. Statistical tables, Greek symbols.
- **Links**. Access to ongoing development of teaching materials associated with the book, including on-line self-assessment.

## Videos

The Website hosts a large number of feedback and instructional videos that have been developed since the first edition of the book was published. Most of these are very short (a few minutes) and provide students with the type of feedback they might expect to receive when asking

a tutor 'how to do' a particular question or computer technique. The videos are targeted to produce support just at the point when the student is really involved with trying to understand a particular detailed problem, and provide the focused help that is both required and very welcome.

These videos are used by students of all abilities: advanced students use them just as a quick check on their own self-study, but weaker students can pause and rerun the videos to provide a very effective self-managed 'tutorial'.

The video formats include a 'hand-written' format for paper-based answers, and 'keystroke' demonstrations for computer-based problems. These match directly the form and content of the knowledge and skills that the student is trying to acquire. The separate videos can be viewed directly and quickly over the Internet, using flash technology which is already loaded with most Internet browsers.

# Contents

# 1

# Mathematics and Statistics in Science

## Overview

Science students encounter mathematics and statistics in three main areas:

- Understanding and using theory.
- Carrying out experiments and analysing results.
- Presenting data in laboratory reports and essays.

Unfortunately, many students do not fully appreciate the need for understanding mathematics and/or statistics until it suddenly confronts them in a lecture or in the write-up of an experiment. There is indeed a 'chicken and egg' aspect to the problem:

> Some science students have little enthusiasm to study mathematics until it appears in a lecture or tutorial – by which time it is too late! Without the mathematics, they cannot *fully* understand the science that is being presented, and they drift into a habit of accepting a 'second-best' science *without* mathematics. The end result could easily be a drop of at least one grade in their final degree qualification.

All science is based on a *quantitative* understanding of the world around us – an understanding described ultimately by *measurable* values. Mathematics and statistics are merely the processes by which we handle these quantitative values in an effective and logical way.

Mathematics and statistics provide the network of links that tie together the details of our understanding, and create a sound basis for a fundamental appreciation of science as a whole. Without these quantifiable links, the ability of science to predict and move forward into new areas of understanding would be totally undermined.

In recent years, the data handling capability of information technology has made mathematical and statistical calculations far easier to perform, and has transformed the day-to-day work in many areas of science. In particular, a good spreadsheet program, like Excel, enables both scientists and students to carry out extensive calculations quickly, and present results and reports in a clear and accurate manner.

*Essential Mathematics and Statistics for Science 2nd Edition* Graham Currell and Antony Dowman

## 1.1 Data and Information

Real-world information is expressed in the mathematical world through **data**.

In science, some data values are believed to be fixed in nature. We refer to values that are fixed as **constants**, e.g. the constant $c$ is often used to represent the speed of light in a vacuum, $c = 3.00 \times 10^8\ \mathrm{m\,s^{-1}}$.

However, most measured values are subject to change. We refer to these values as **variables**, e.g. $T$ for temperature, pH for acidity.

The term **parameter** refers to a variable that can be used to describe a relevant characteristic of a scientific system, or a statistical population (see 7.2.2), e.g. the actual pH of a buffer solution, or the average (mean) age of the whole UK population. The term **statistic** refers to a variable that is used to describe a relevant characteristic of a *sampled* (see 7.2.2) set of data, e.g. five repeated measurements of the concentration of a solution, or the average (mean) age of 1000 members of the UK population.

Within this book we use the convention of printing letters and symbols that represent quantities (constants and variables) in italics, e.g. $c$, $T$ and $p$.

The letters that represent units are presented in normal form, e.g. $\mathrm{m\,s^{-1}}$ gives the units of speed in metres per second.

There is an important relationship between data and information, which appears when analysing more complex data sets. It is a basic rule that:

> It is impossible to get more 'bits' of information from a calculation than the number of 'bits' of data that is put into the calculation.

For example, if a chemical mixture contains three separate compounds, then it is necessary to make at least three separate measurements on that mixture before it is possible to calculate the concentration of each separate compound.

In mathematics and statistics, the *number* of bits of information that are available in a data set is called the **degrees of freedom**, $df$, of that data set. This value appears in many statistical calculations, and it is usually easy to calculate the number of degrees of freedom appropriate to any given situation.

## 1.2 Experimental Variation and Uncertainty

The uncertainty inherent in scientific information is an important theme that appears throughout the book.

The **true value** of a variable is the value that we would measure if our measurement process were 'perfect'. However, because no process is perfect, the 'true value' is not normally known.

The **observed value** is the value that we produce as our *best estimate* of the true value.

The **error** in the measurement is the difference between the true value and the observed value:

$$\text{Error} = \text{Observed value} - \text{True value} \qquad [1.1]$$

As we do not normally know the 'true value', we cannot therefore know the actual error in any particular measurement. However, it is important that we have some idea of how large the error might be.

The **uncertainty** in the measurement is our *best estimate* of the magnitude of possible errors. The magnitude of the uncertainty must be derived on the basis of a proper understanding of the measurement process involved and the system being measured. The statistical interpretation of uncertainty is derived in 8.2.

The uncertainty in experimental measurements can be divided into two main categories:

**Measurement uncertainty.** Variations in the actual process of measurement will give some differences when the same measurement is repeated under exactly the same conditions. For example, repeating a measurement of alcohol level in the same blood sample may give results that differ by a few milligrams in each 100 millilitres of blood.

**Subject uncertainty.** A subject is a representative example of the system (9.1) being measured, but many of the systems in the real world have inherent variability in their responses. For example, in testing the effectiveness of a new drug, every person (subject) will have a slightly different reaction to that drug, and it would be necessary to carry out the test on a wide range of people before being confident about the 'average' response.

Whatever the source of uncertainty, it is important that any experiment must be designed both to counteract the effects of uncertainty and to quantify the magnitude of that uncertainty.

Within each of the two types of uncertainty, *measurement* and *subject*, it is possible to identify two further categories:

**Random error.** Each subsequent measurement has a random error, leading to *imprecision* in the result. A measurement with a low random error is said to be a *precise* measurement.

**Systematic error.** Each subsequent measurement has the same recurring error. A systematic error shows that the measurement is *biased*, e.g. when setting the liquid level in a burette, a particular student may always set the meniscus of the liquid a little too low.

The **precision** of a measurement is the best estimate for the purely *random error* in a measurement.

The **trueness** of a measurement is the best estimate for the *bias* in a measurement.

The **accuracy** of a measurement is the best estimate for the *overall error* in the final result, and includes both the effects of a lack of precision (due to random errors) and bias (due to systematic errors).

### Example 1.1

Four groups of students each measure the pH (acidity) of a sample of soil, with each group preparing five replicate samples for testing. The results are given in Figure 1.1.

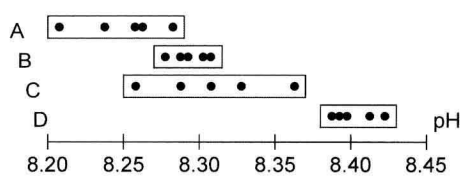



**Figure 1.1** Precision and bias in experimental data.

What can be said about the *accuracy* of their results?

It is possible to say that the results from groups A and C show greater *random uncertainty* (less precision) than groups B and D. This could be due to such factors as a lack of care in preparing the five samples for testing, or some electronic instability in the pH meter being used.

Groups B and D show greater precision, but at least one of B or D must have some *bias* in their measurements, i.e. poor 'trueness'. The bias could be due to an error in setting the pH meter with a buffer solution, which would then make every one of the five measurements in the set wrong by the same amount.

With the information given, very little can be said about the overall *accuracy* of the measurements; the 'true' value is not known, and there is no information about possible bias in any of the results. For example *if the true value were* $pH = 8.40$, this would mean that groups A, B and C were all biased, with the most *accurate* measurement being group D.

The effect of random errors can be managed and quantified using suitable statistical methods (8.2, 8.3 and 15.1.2). The presentation of uncertainty as *error bars* on graphs is developed in an Excel tutorial on the Website.

Systematic errors are more difficult to manage in an experiment, but good experiment design (Chapter 15) aims to counteract their effect as much as possible.

## 1.3 Mathematical Models in Science

A fundamental building block of both science and mathematics is the *equation*.

Science uses the equation as a *mathematical model* to define the *relationship* between one or more factors in the real world (3.1.6). It may then be possible to use mathematics to investigate how that equation may lead to *new conclusions* about the world.

Perhaps the most famous equation, arising from the general theory of relativity, is:

$$E = mc^2$$

which relates the amount of energy, $E$ (J), that would be released if a mass, $m$ (kg), of matter was converted into energy (e.g. in a nuclear reactor). $E$ and $m$ are both variables and the constant $c(= 3.00 \times 10^8\ \mathrm{m\,s^{-1}})$ is the speed of light.

**Example 1.2**

Calculate the amount of matter, $m$, that must be converted *completely* into energy, if the amount of energy, $E$, is equivalent to that produced by a medium-sized power station in one year: $E = 1.8 \times 10^{13}$ J.