Contents

Preface	xi
Acknowledgement	xvii
1 Modelling the Changing Arctic Ice	1
1.1 Global Climate Change and Our First Model	1
1.1.1 Introduction and motivation	1
1.1.2 A first look at some data and the importance of units	1
1.1.3 Using variables for our data and summarizing the data with tables	3
1.1.4 Plotting our data for a visual summary	4
1.1.5 Making assumptions about physical processes	5
1.1.6 A first model using equations and making predictions	7
1.1.7 Comparing data and models with plots and lines of best-fit	10
1.2 The Linear Model	12
1.2.1 Using parameters and estimating unknowns	12
1.2.2 Using residuals and the sum of squared residuals to measure model accuracy	14
1.2.3 The constant model	17
1.2.4 The zero-intercept model	23
1.2.5 Using transformations to simplify data and model building	25
1.2.6 Using transformations to fit the linear model	30
1.2.7 Fitting the linear model in practice: variance, covariance and trend-lines	33
1.2.8 The best-fit linear model and extrapolations	35
1.3 Beyond Linear Models	36
1.3.1 Quadratic models	36
1.3.2 Polynomial models	39
1.4 Conclusions and Key Results	42
1.5 Resources for Further Study	42
2 Modelling Bacterial Population Growth	44
2.1 Rapid Growth of Microbial Populations	44
2.1.1 Salmonella and Ebola	44
2.1.2 Setting out the problem I: choosing our variables and units	46
2.1.3 Setting out the problem II: data visualization	48
2.2 The Exponential Model	49
2.2.1 The growth of bacterial populations: the physical process and idealizations	49
2.2.2 The growth of bacterial populations: a simple model	52
2.2.3 The exponential model	54

2.2.4 Exponential curves 56
2.2.5 Applying the exponential model to data 59
2.2.6 Simplifying the problem using transformations of data: exponentiation and logarithms 62
2.2.7 Building an exponential model of population growth: log-plots and linear fits 66
2.2.8 Extrapolating and checking our physical assumptions 70
2.3 Difference Equations 71
2.3.1 Stating modelling assumptions with difference equations and the constant model 71
2.3.2 A difference equation for the linear model 73
2.3.3 A difference equation for the exponential model 75
2.3.4 Initial values for difference equations and discrete versus continuous models 77
2.4 Conclusions and Key Results 79
2.5 Resources for Further Study 80

3 Modelling the Growth of Human Populations 82
3.1 The Growth of the Total Human Population 82
3.1.1 The importance of predicting the size of the total human population 82
3.1.2 Setting out the problem 82
3.1.3 Using functions to transform data 84
3.1.4 Transformations and inverse functions 86
3.1.5 The exponential model revisited I: using functions to transform our data 87
3.1.6 The exponential model revisited II: using functions to transform our models 89
3.2 The Exponential-Quadratic Model 92
3.2.1 Late population growth: the exponential model 92
3.2.2 A quadratic model for log-populations 95
3.2.3 Comparing the exponential and exponential-quadratic models 96
3.2.4 Extrapolating with the exponential and exponential-quadratic models 98
3.3 Quadratic Difference Equations and Population Models 99
3.3.1 Developing difference equations with data and fitting 99
3.3.2 Late time population growth: a quadratic difference equation 103
3.3.3 Applying difference equations as recurrence relations 105
3.3.4 Comparing the long-term behaviour of our models 107
3.4 Conclusions and Key Results 109
3.5 Resources for Further Study 110

4 Modelling Radioactive Decay to Determine the Age of the Earth 111
4.1 Introduction to Chronological Dating 111
4.1.1 Chronological dating, human history and the geological timescale 111

4.1.2 Models as functions and inverting models 113
4.1.3 The change in atomic populations through radioactive decay: fractional populations 114
4.2 The Exponential Model Applied to Decay 117
4.2.1 The radioactive decay of atoms 117
4.2.2 The exponential decay of populations: bacterial death and radioactive decay 119
4.2.3 The exponential model base-e: irrational numbers and the exponential function 122
4.2.4 An exponential model for ^{224}Ra: fitting and interpreting the parameters 123
4.2.5 Predicting ages with fractional populations 127
4.3 Differential Equations 128
4.3.1 Instantaneous speeds and average speeds 128
4.3.2 Approximating derivatives with differences 130
4.3.3 Defining derivatives with differences and calculating them with differentiation 133
4.3.4 Differentiation rules for common functions 135
4.3.5 The idea of differential equations and their approximation with difference equations 137
4.3.6 Predicting ages with populations: multiple populations 139
4.3.7 Predicting the age of the Earth from zircon 144
4.4 Conclusions and Key Results 144
4.5 Resources for Further Study 145

5 Modelling the Distribution of Butterfly Species 147
5.1 Improving Measures of the Goodness of Fit 147
5.1.1 Conservation and distribution of species 147
5.1.2 Species–area relationships for butterflies 148
5.1.3 Measuring goodness of fit with a single value: the coefficient of determination R^2 150
5.1.4 Calculating the baseline SSR with variance 153
5.1.5 Calculating R^2 with Pearson's correlation coefficient and covariance 154
5.2 The Power-Law Model 155
5.2.1 Using R^2 with transformations for model development 155
5.2.2 Linear models for log-log data and power-laws 159
5.2.3 Power-law curves 163
5.2.4 Understanding our failures: using residuals for error analysis 163
5.3 Models with Multiple Predictors 165
5.3.1 Data with multiple predictors: representation of data and using categories 165
5.3.2 Using multiple features for prediction 167
5.3.3 Plotting models with multiple features 170
5.3.4 The multiple linear model and interactions 170
5.3.5 Writing the polynomial model as a multiple linear model 173
5.3.6 Linear least-squares problems 175

5.3.7 A general framework for models with interactions 177
5.3.8 The problem of overfitting and increasing R^2 178
5.3.9 Fitting with R software and using adjusted R-squared 179
5.3.10 Simplifying models by setting parameters to zero 181
5.3.11 Errors in data and standard errors of best-fit values 183
5.3.12 Using statistical hypothesis testing to simplify models 188
5.4 Conclusions and Key Results 191
5.5 Resources for Further Study 191

6 Volcanic Gas, Environmental Damage and Statistical Models 193
6.1 Introduction: A Binary Response Model for Vegetation Damage 193
6.1.1 Natural environmental damage 193
6.1.2 A first look at some data 193
6.1.3 Binary classification: a model for predicting vegetation damage from gas concentration 195
6.2 Building a Simple Model for Classification: The Linear Model and a Step Function 200
6.2.1 A linear model for concentration and distance 200
6.2.2 Fitting the constant model and minimization of functions 200
6.2.3 Minimization of functions using differentiation 203
6.2.4 Fitting the constant model using differentiation 206
6.2.5 Finding the SSR of the linear model 206
6.2.6 Simplifying the SSR with summation rules 208
6.2.7 Minimizing functions with multiple inputs: *partial derivatives* 211
6.2.8 Minimizing functions with multiple inputs: simultaneous equations 214
6.2.9 Fitting the linear model 215
6.2.10 *Predicting vegetation index from distance* 218
6.2.11 The danger of predictions made with certainty 219
6.3 Statistical Models 222
6.3.1 The ideas of probability and statistical models 222
6.3.2 *A statistical model for the vegetation index and SO_2 concentration* 224
6.3.3 Building an improved statistical model with a sigmoid function 226
6.3.4 Comparing statistical models to data with likelihoods and log-likelihoods 228
6.3.5 A simple statistical model for the concentration and distance 230
6.3.6 Using Gaussian functions for statistical models 231
6.3.7 Probability density functions 235
6.3.8 Maximum likelihood estimation 239
6.4 Conclusions and Key Results 245
6.5 Resources for Further Study 246

Appendices 247
A1. Modelling Toolbox 247
A2. Exponent Laws 249

A3. Logarithmic Identities 249
A4. Differentiation Rules 250
A5. Using Common Inbuilt Functions 251
A6. List of Parameter Values and Results to 10 Significant Figures 251

References and Further Reading 255

Index 259