How to be a Quantitative Ecologist

The A to R of green mathematics & statistics

How I chose to write this book, and why you might choose to read it

Preface

0. How to start a meaningful relationship with your computer

Introduction to R

0.1 What is R? 1 0.8 Basic R usage 8
0.2 Why use R for this book? 2 0.9 Importing data from a spreadsheet 9
0.3 Computing with a scientific package like R 2 0.10 Storing data in data frames 10
0.4 Installing and interacting with R 3 0.11 Exporting data from R 12
0.5 Style conventions 5 0.12 Quitting R 12
0.6 Valuable R accessories 6 Further reading 13
0.7 Getting help 7 References 13
## 1. How to make mathematical statements

### Numbers, equations and functions

<table>
<thead>
<tr>
<th>Number</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Qualitative and quantitative scales</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>• Habitat classifications</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>Numbers</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>• Observations of spatial abundance</td>
<td></td>
</tr>
<tr>
<td>1.3</td>
<td>Symbols</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>• Population size and carrying capacity</td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>Logical operations</td>
<td>21</td>
</tr>
<tr>
<td>1.5</td>
<td>Algebraic operations</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>• Size matters in male garter snakes</td>
<td></td>
</tr>
<tr>
<td>1.6</td>
<td>Manipulating numbers</td>
<td>25</td>
</tr>
<tr>
<td>1.7</td>
<td>Manipulating units</td>
<td>26</td>
</tr>
<tr>
<td>1.8</td>
<td>Manipulating expressions</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>• Energy acquisition in voles</td>
<td></td>
</tr>
<tr>
<td>1.9</td>
<td>Polynomials</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>• The law of mass action in epidemiology</td>
<td></td>
</tr>
<tr>
<td>1.10</td>
<td>Equations</td>
<td>32</td>
</tr>
<tr>
<td>1.11</td>
<td>First order polynomial equations</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>• Population size and composition</td>
<td></td>
</tr>
<tr>
<td>1.12</td>
<td>Proportionality and scaling: a special kind of first order polynomial equation</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>• Simple mark-recapture</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Converting density to population size</td>
<td></td>
</tr>
<tr>
<td>1.13</td>
<td>Second and higher order polynomial equations</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>• Estimating the number of infected animals from the rate of infection</td>
<td></td>
</tr>
<tr>
<td>1.14</td>
<td>Systems of polynomial equations</td>
<td>38</td>
</tr>
<tr>
<td>1.15</td>
<td>Inequalities</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>• Deriving population structure from data on population size</td>
<td></td>
</tr>
<tr>
<td>1.16</td>
<td>Coordinate systems</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>• Minimum energetic requirements in voles</td>
<td></td>
</tr>
<tr>
<td>1.17</td>
<td>Complex numbers</td>
<td>44</td>
</tr>
<tr>
<td>1.18</td>
<td>Relations and functions</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>• Food webs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Mating systems in animals</td>
<td></td>
</tr>
<tr>
<td>1.19</td>
<td>The graph of a function</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>• Two aspects of vole energetics</td>
<td></td>
</tr>
<tr>
<td>1.20</td>
<td>First order polynomial functions</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>• Population stability in a time series</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Population stability and population change</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Visualising goodness-of-fit</td>
<td></td>
</tr>
<tr>
<td>1.21</td>
<td>Higher order polynomial functions</td>
<td>55</td>
</tr>
<tr>
<td>1.22</td>
<td>The relationship between equations and functions</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>• Extent of an epidemic when the transmission rate exceeds a critical value</td>
<td></td>
</tr>
<tr>
<td>1.23</td>
<td>Other useful functions</td>
<td>58</td>
</tr>
<tr>
<td>1.24</td>
<td>Inverse functions</td>
<td>60</td>
</tr>
<tr>
<td>1.25</td>
<td>Functions of more than one variable</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>• Two aspects of vole energetics</td>
<td></td>
</tr>
<tr>
<td>Further reading</td>
<td></td>
<td>64</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>65</td>
</tr>
</tbody>
</table>
CONTENTS

2. How to describe regular shapes and patterns 67

Geometry and trigonometry

2.1 Primitive elements 68

2.2 Axioms of Euclidean geometry 69

• Suicidal lemmings, parsimony, evidence and proof

2.3 Propositions 70

• Radio-tracking of terrestrial animals

2.4 Distance between two points 76

• Spatial autocorrelation in ecological variables

2.5 Areas and volumes 78

• Hexagonal territories

2.6 Measuring angles 80

• The bearing of a moving animal

2.7 The trigonometric circle 82

• The position of a seed following dispersal

2.8 Trigonometric functions 83

2.9 Polar coordinates 85

• Random walks

2.10 Graphs of trigonometric functions 86

2.11 Trigonometric identities 88

• A two-step random walk

2.12 Inverses of trigonometric functions 89

• Displacement during a random walk

2.13 Trigonometric equations 91

• VHF tracking for terrestrial animals

2.14 Modifying the basic trigonometric graphs 93

• Nocturnal flowering in dry climates

2.15 Superimposing trigonometric functions 96

• More realistic model of nocturnal flowering

2.16 Spectral analysis 99

• Dominant frequencies in density fluctuations of Norwegian lemming populations

• Spectral analysis of oceanographic covariates

2.17 Fractal geometry 102

• Availability of coastal habitat

• Fractal dimension of the Koch curve

Further reading 105

References 106

3. How to change things, one step at a time 107

Sequences, difference equations and logarithms

3.1 Sequences 108

• Reproductive output in social wasps

• Unrestricted population growth

3.2 Difference equations 111

• More realistic models of population growth

3.3 Higher order difference equations 114

• Delay-difference equations in a biennial plant

3.4 Initial conditions and parameters 115

3.5 Solutions of a difference equation 116

3.6 Equilibrium solutions 117

• Harvesting an unconstrained population

• Visualising the equilibria
## CONTENTS

3.7 Stable and unstable equilibria  
- Parameter sensitivity and ineffective fishing quotas  
- Stable and unstable equilibria in a density-dependent population

3.8 Investigating stability  
- Cobweb plot for an unconstrained, harvested population  
- Conditions for stability under unrestricted growth

3.9 Chaos  
- Chaos in a model with density dependence

3.10 Exponential function  
- Modelling bacterial loads in continuous time  
- A negative blue tit? Using exponential functions to constrain models

3.11 Logarithmic function  
- Log-transforming population time series

3.12 Logarithmic equations

Further reading

References

## 4. How to change things, continuously 137

**Derivatives and their applications**

4.1 Average rate of change  
- Seasonal tree growth  
- Tree growth

4.2 Instantaneous rate of change  
- Methane concentration around termite mounds

4.3 Limits  
- Plotting change in tree biomass  
- Linear tree growth

4.4 The derivative of a function  
- Spatial gradients

4.5 Differentiating polynomials  
- Consumption rates of specialist predators

4.6 Differentiating other functions  
- The slope of the sea-floor

4.7 The chain rule

4.8 Higher order derivatives  
- Maximum rate of disease transmission  
- The marginal value theorem

4.9 Derivatives of functions of many variables  
- Unconstrained population growth  
- Density dependence and proportional harvesting

4.10 Optimisation  
- Series expansions

4.11 Local stability for difference equations  
- Further reading

4.12 Series expansions

References
## 5. How to work with accumulated change

*Integrals and their applications*

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Antiderivatives</td>
<td>178</td>
<td>• Total number of arrivals and departures at migratory stop-over&lt;br&gt;• Invasion fronts&lt;br&gt;• Diving in seals</td>
</tr>
<tr>
<td>5.2 Indefinite integrals</td>
<td>181</td>
<td>• Failing to stop invasion fronts&lt;br&gt;• Allometry</td>
</tr>
<tr>
<td>5.3 Three analytical methods of integration</td>
<td>182</td>
<td>• A differential equation for a plant invasion front&lt;br&gt;• Stopping invasion fronts</td>
</tr>
<tr>
<td>5.4 Summation</td>
<td>187</td>
<td>• Exponential population growth in continuous time&lt;br&gt;• Metapopulations</td>
</tr>
<tr>
<td>5.5 Area under a curve</td>
<td>190</td>
<td>• Constrained growth in continuous time&lt;br&gt;• Swimming speed in seals</td>
</tr>
<tr>
<td>5.6 Definite integrals</td>
<td>193</td>
<td>• Constrained growth in continuous time&lt;br&gt;• Swimming speed in seals</td>
</tr>
<tr>
<td>5.7 Some properties of definite integrals</td>
<td>195</td>
<td>• The Levins model for metapopulations&lt;br&gt;• Total reproductive output in social wasps&lt;br&gt;• Net change in number of birds at migratory stop-over</td>
</tr>
</tbody>
</table>

Further reading

References | 212 | |

## 6. How to keep stuff organised in tables

*Matrices and their applications*

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Matrices</td>
<td>214</td>
<td>• Plant community composition&lt;br&gt;• Inferring diet from fatty acid analysis</td>
</tr>
<tr>
<td>6.2 Matrix operations</td>
<td>217</td>
<td>• Stage-structured seal populations&lt;br&gt;• Movement in metapopulations&lt;br&gt;• Equilibrium of linear Leslie model</td>
</tr>
<tr>
<td>6.3 Geometric interpretation of vectors and square matrices</td>
<td>221</td>
<td>• Stability in a linear Leslie model&lt;br&gt;• Random walks as sequences of vectors</td>
</tr>
<tr>
<td>6.4 Solving systems of equations with matrices</td>
<td>223</td>
<td>• Stable age structure in a linear Leslie model&lt;br&gt;• Plant community composition</td>
</tr>
<tr>
<td>6.5 Markov chains</td>
<td>227</td>
<td>• Growth in patchy populations&lt;br&gt;• Redistribution between population patches</td>
</tr>
<tr>
<td>6.6 Eigenvalues and eigenvectors</td>
<td>230</td>
<td>• Metapopulation growth&lt;br&gt;• A fragmented population in continuous time</td>
</tr>
<tr>
<td>6.7 Leslie matrix models</td>
<td>233</td>
<td>• Phase-space for a two-patch metapopulation&lt;br&gt;• Stage-structured seal populations&lt;br&gt;• Equilibrium of linear Leslie model</td>
</tr>
<tr>
<td>6.8 Analysis of linear dynamical systems</td>
<td>237</td>
<td>• Stability analysis of a two-patch metapopulation</td>
</tr>
</tbody>
</table>
6.9 Analysis of nonlinear dynamical systems 243
- The Lotka–Volterra, predator–prey model
- Stability analysis of the Lotka–Volterra model

Further reading 249
References 249

7. How to visualise and summarise data 251
Descriptive statistics

7.1 Overview of statistics 252
- 7.8 Skewness and kurtosis 266
- 7.9 Graphical summaries 267
- 7.10 Data sets with more than one variable 268
- 7.11 Association between qualitative variables 268
- 7.12 Association between quantitative variables 270
- 7.13 Joint frequency distributions 273
- Further reading 276
- References 276

7.2 Statistical variables 253
- Activity budgets in honey bees
- 7.5 Frequency distributions 256
- Activity budgets from different studies
- Visualising activity budgets
- Height of tree ferns
- Gannets on Bass rock

7.3 Populations and samples 255
- Production of gannet chicks

7.4 Single-variable samples 255

7.5 Frequency distributions 256
- Activity budgets in honey bees
- Activity budgets from different studies
- Visualising activity budgets
- Height of tree ferns
- Gannets on Bass rock

7.6 Measures of centrality 260
- Chick rearing in red grouse
- Swimming speed in grey seals
- Median of chicks reared by red grouse

7.7 Measures of spread 263
- Gannet foraging

8. How to put a value on uncertainty 279
Probability

8.1 Random experiments and event spaces 280
- Assumptions of random experiments

8.2 Events 281
- Plant occurrence in survey quadrats
- Overlapping events
- Mutually exclusive events

8.3 Frequentist probability 284
- Fluctuating frequency of newborn male wildebeest

8.4 Equally likely events 285
- Something is certain to occur (even if it is nothing)
- Undirected movement

8.5 The union of events 286
- Seed germination on a gridded landscape
- Coexisting sparrows

8.6 Conditional probability 288
- Territoriality and survival in red grouse
8.7 Independent events
- Sex of successive calves

8.8 Total probability
- Seed germination in a heterogeneous environment

8.9 Bayesian probability
- Does the sex ratio at birth deviate from 1:1?

Further reading
References

9. How to identify different kinds of randomness

9.1 Probability distributions
- Probability distribution of nominal variables

9.2 Discrete probability distributions
- Beetle eggs per cluster: a count
- PMF for egg cluster size
- CDF for egg cluster size

9.3 Continuous probability distributions
- Are all fern heights equally likely?
- CDF for tree fern heights
- PDF for tree fern heights

9.4 Expectation
- Mean egg cluster size
- Expected tree fern heights

9.5 Named distributions

9.6 Equally likely events: the uniform distribution
- River otter home ranges

9.7 Hit or miss: the Bernoulli distribution
- Bernoulli births

9.8 Count of occurrences in a given number of trials: the binomial distribution
- Wildebeest reproductive histories
- Wildebeest reproductive output

9.9 Counting different types of occurrences: the multinomial distribution
- Metapopulation transitions

9.10 Number of occurrences in a unit of time or space: the Poisson distribution
- Data on plant abundance

9.11 The gentle art of waiting: geometric, negative binomial, exponential and gamma distributions
- Catastrophic events and species extinctions

9.12 Assigning probabilities to probabilities: the beta and Dirichlet distributions

9.13 Perfect symmetry: the normal distribution
- Weight distribution in voles

9.14 Because it looks right: using probability distributions empirically
- A beta prior for wildebeest sex at-birth

9.15 Mixtures, outliers and the t-distribution
- Bimodal weight distribution in voles

9.16 Joint, conditional and marginal probability distributions
- Feeding site fidelity in kittiwakes

9.17 The bivariate normal distribution
- Height and root depth of tree ferns

9.18 Sums of random variables: the central limit theorem
- Food provisioning in starlings
- Mixed-diet provisioning in starlings
10. How to see the forest from the trees 345

Estimation and testing

10.1 Estimators and their properties 345
- Sampling gannet egg weights
- Maximum likelihood estimator for wildebeest sex ratio

10.2 Normal theory 347
- Bayesian estimator for wildebeest sex ratio

10.3 Estimating the population mean 348
- Bayesian estimator for wildebeest sex ratio

10.4 Estimating the variance of a normal population 349
- Comparing ML and Bayes estimates of sex ratio at birth

10.5 Confidence intervals 350
- Estimating the unknown parameters of the gannet egg weight distribution

10.6 Inference by bootstrapping 354
- Bootstrap inference for gannet age distribution

10.7 More general estimation methods 356
- Estimating parameters for animal movement

10.8 Estimation by least squares 358
- Weight distribution in voles of different ages
- Mean weight in a cohort of voles

10.9 Estimation by maximum likelihood 359
- Maximum likelihood estimation of mean and variance of vole weight distribution

10.10 Bayesian estimation 364
- Maximum likelihood estimator for wildebeest sex ratio

10.11 Link between maximum likelihood and Bayesian estimation 368
- Comparing ML and Bayes estimates of sex ratio at birth

10.12 Hypothesis testing: rationale 369
- Paired samples of gannet condition
- How to rank observations

10.13 Tests for the population mean 370
- Tree selection in Peruvian ants

10.14 Tests comparing two different means 373
- Niche partitioning in Peruvian ants

10.15 Hypotheses about qualitative data 376
- How to rank observations

10.16 Hypothesis testing debunked 379

Further reading 379

References 379
11. How to separate the signal from the noise

Statistical modelling

11.1 Comparing the means of several populations 382
- Samples of gannet condition
- Estimates of multiple population averages

11.2 Simple linear regression 386
- Density dependence of gannet condition
- Modelling density dependence of gannet condition

11.3 Prediction 390
- Predicting density-dependent changes in gannet condition

11.4 How good is the best-fit line? 391
- Diagnostics for the density dependence model

11.5 Multiple linear regression 394
- Combined effects of density and the environment

11.6 Model selection 396
- Should the beating of a butterfly’s wings be used to explain gannet condition?
- Model selection by adjusted $r^2$ and AIC
- Collinearity in covariates of gannet condition

11.7 Generalised linear models 400
- Many response variables are constrained
- Log-transforming fecundity data
- Modelling count data

11.8 Evaluation, diagnostics and model selection for GLMs 407
- Seeing the linear model as a special case of the GLM
- Likelihood for a log-linear GLM
- Modelling senescence in turtles
- Estimating survival as a function of age

11.9 Modelling dispersion 409

11.10 Fitting more complicated models to data: polynomials, interactions, nonlinear regression 410
- Density dependence and polynomial terms
- Predation, immigration and interaction terms

11.11 Letting the data suggest more complicated models: smoothing 415
- Distribution along a linear habitat
- Otter density estimation by kernel smoothing
- Otter density estimation by GAM

11.12 Partitioning variation: mixed effects models 419
- Samples of gannet condition
- Modelling individual and colony variation
- Why not use colony as a factor?

Further reading 422

References 423

12. How to measure similarity

Multivariate methods

12.1 The problem with multivariate data 426
- Characterising environmental similarity
- Characterising patterns of occurrence

12.2 Ordination in general 427
- Correlations represent redundancy

12.3 Principal components analysis 428
- Collinearities between four environmental variables
- PCA for four environmental variables

12.4 Clustering in general 432
Identifying functional groups in ecological communities
Clustering data frame for Antarctic species

12.5 Agglomerative hierarchical clustering
Dendrogram for Antarctic species

12.6 Nonhierarchical clustering: k means analysis

12.7 Classification in general

12.8 Logistic regression: two classes
Characterising fern habitat

12.9 Logistic regression: many classes
Classification of whale vocalisations

Further reading
References

Appendix: Formulae

R Index

Index