

Contents

Preface.....	xv
Authors.....	xix
1. Introduction.....	1
1.1 Different Types of Scientific Study.....	1
1.2 Relating Sample Results to More General Populations.....	3
1.3 Constructing Models to Represent Reality.....	4
1.4 Using Linear Models.....	7
1.5 Estimating the Parameters of Linear Models.....	8
1.6 Summarizing the Importance of Model Terms.....	9
1.7 The Scope of This Book.....	11
2. A Review of Basic Statistics.....	13
2.1 Summary Statistics and Notation for Sample Data.....	13
2.2 Statistical Distributions for Populations.....	16
2.2.1 Discrete Data.....	17
2.2.2 Continuous Data.....	22
2.2.3 The Normal Distribution.....	24
2.2.4 Distributions Derived from Functions of Normal Random Variables.....	26
2.3 From Sample Data to Conclusions about the Population.....	28
2.3.1 Estimating Population Parameters Using Summary Statistics.....	28
2.3.2 Asking Questions about the Data: Hypothesis Testing.....	29
2.4 Simple Tests for Population Means.....	30
2.4.1 Assessing the Mean Response: The One-Sample t-Test.....	30
2.4.2 Comparing Mean Responses: The Two-Sample t-Test.....	32
2.5 Assessing the Association between Variables.....	36
2.6 Presenting Numerical Results.....	39
Exercises.....	41
3. Principles for Designing Experiments.....	43
3.1 Key Principles.....	43
3.1.1 Replication.....	46
3.1.2 Randomization.....	48
3.1.3 Blocking.....	51
3.2 Forms of Experimental Structure.....	52
3.3 Common Forms of Design for Experiments.....	57
3.3.1 The Completely Randomized Design.....	57
3.3.2 The Randomized Complete Block Design.....	58
3.3.3 The Latin Square Design.....	59
3.3.4 The Split-Plot Design.....	60
3.3.5 The Balanced Incomplete Block Design.....	61
3.3.6 Generating a Randomized Design.....	62
Exercises.....	62

4. Models for a Single Factor	69
4.1 Defining the Model	69
4.2 Estimating the Model Parameters	73
4.3 Summarizing the Importance of Model Terms	74
4.3.1 Calculating Sums of Squares	76
4.3.2 Calculating Degrees of Freedom and Mean Squares	80
4.3.3 Calculating Variance Ratios as Test Statistics	81
4.3.4 The Summary ANOVA Table	82
4.4 Evaluating the Response to Treatments	84
4.4.1 Prediction of Treatment Means	84
4.4.2 Comparison of Treatment Means	85
4.5 Alternative Forms of the Model	88
Exercises	90
5. Checking Model Assumptions	93
5.1 Estimating Deviations	93
5.1.1 Simple Residuals	94
5.1.2 Standardized Residuals	95
5.2 Using Graphical Tools to Diagnose Problems	96
5.2.1 Assessing Homogeneity of Variances	96
5.2.2 Assessing Independence	98
5.2.3 Assessing Normality	101
5.2.4 Using Permutation Tests Where Assumptions Fail	102
5.2.5 The Impact of Sample Size	103
5.3 Using Formal Tests to Diagnose Problems	104
5.4 Identifying Inconsistent Observations	108
Exercises	110
6. Transformations of the Response	113
6.1 Why Do We Need to Transform the Response?	113
6.2 Some Useful Transformations	114
6.2.1 Logarithms	114
6.2.2 Square Roots	119
6.2.3 Logits	120
6.2.4 Other Transformations	121
6.3 Interpreting the Results after Transformation	122
6.4 Interpretation for Log-Transformed Responses	123
6.5 Other Approaches	126
Exercises	127
7. Models with a Simple Blocking Structure	129
7.1 Defining the Model	130
7.2 Estimating the Model Parameters	132
7.3 Summarizing the Importance of Model Terms	134
7.4 Evaluating the Response to Treatments	140
7.5 Incorporating Strata: The Multi-Stratum Analysis of Variance	141
Exercises	146

8. Extracting Information about Treatments	149
8.1 From Scientific Questions to the Treatment Structure	150
8.2 A Crossed Treatment Structure with Two Factors	152
8.2.1 Models for a Crossed Treatment Structure with Two Factors.....	153
8.2.2 Estimating the Model Parameters	155
8.2.3 Assessing the Importance of Individual Model Terms	158
8.2.4 Evaluating the Response to Treatments: Predictions from the Fitted Model.....	160
8.2.5 The Advantages of Factorial Structure	162
8.2.6 Understanding Different Parameterizations	163
8.3 Crossed Treatment Structures with Three or More Factors	164
8.3.1 Assessing the Importance of Individual Model Terms	166
8.3.2 Evaluating the Response to Treatments: Predictions from the Fitted Model	171
8.4 Models for Nested Treatment Structures	173
8.5 Adding Controls or Standards to a Set of Treatments	179
8.6 Investigating Specific Treatment Comparisons	182
8.7 Modelling Patterns for Quantitative Treatments	190
8.8 Making Treatment Comparisons from Predicted Means	195
8.8.1 The Bonferroni Correction	196
8.8.2 The False Discovery Rate	197
8.8.3 All Pairwise Comparisons.....	198
8.8.3.1 The LSD and Fisher's Protected LSD.....	198
8.8.3.2 Multiple Range Tests.....	199
8.8.3.3 Tukey's Simultaneous Confidence Intervals	200
8.8.4 Comparison of Treatments against a Control.....	201
8.8.5 Evaluation of a Set of Pre-Planned Comparisons	201
8.8.6 Summary of Issues	205
Exercises	206
9. Models with More Complex Blocking Structure	209
9.1 The Latin Square Design.....	209
9.1.1 Defining the Model.....	211
9.1.2 Estimating the Model Parameters	211
9.1.3 Assessing the Importance of Individual Model Terms	212
9.1.4 Evaluating the Response to Treatments: Predictions from the Fitted Model.....	215
9.1.5 Constraints and Extensions of the Latin Square Design	217
9.2 The Split-Plot Design	220
9.2.1 Defining the Model.....	222
9.2.2 Assessing the Importance of Individual Model Terms	223
9.2.3 Evaluating the Response to Treatments: Predictions from the Fitted Model.....	225
9.2.4 Drawbacks and Variations of the Split-Plot Design.....	228
9.3 The Balanced Incomplete Block Design.....	232
9.3.1 Defining the Model.....	235
9.3.2 Assessing the Importance of Individual Model Terms	236

9.3.3 Drawbacks and Variations of the Balanced Incomplete Block Design.....	237
Exercises	238
10. Replication and Power.....	241
10.1 Simple Methods for Determining Replication.....	242
10.1.1 Calculations Based on the LSD.....	242
10.1.2 Calculations Based on the Coefficient of Variation	243
10.1.3 Unequal Replication and Models with Blocking	244
10.2 Estimating the Background Variation	245
10.3 Assessing the Power of a Design	245
10.4 Constructing a Design for a Particular Experiment	249
10.5 A Different Hypothesis: Testing for Equivalence.....	253
Exercise	256
11. Dealing with Non-Orthogonality	257
11.1 The Benefits of Orthogonality	257
11.2 Fitting Models with Non-Orthogonal Terms.....	259
11.2.1 Parameterizing Models for Two Non-Orthogonal Factors	259
11.2.2 Assessing the Importance of Non-Orthogonal Terms: The Sequential ANOVA Table	265
11.2.3 Calculating the Impact of Model Terms	269
11.2.4 Selecting the Best Model.....	270
11.2.5 Evaluating the Response to Treatments: Predictions from the Fitted Model.....	270
11.3 Designs with Planned Non-Orthogonality	272
11.3.1 Fractional Factorial Designs	273
11.3.2 Factorial Designs with Confounding.....	274
11.4 The Consequences of Missing Data	274
11.5 Incorporating the Effects of Unplanned Factors	277
11.6 Analysis Approaches for Non-Orthogonal Designs.....	280
11.6.1 A Simple Approach: The Intra-Block Analysis.....	281
Exercises	284
12. Models for a Single Variate: Simple Linear Regression.....	287
12.1 Defining the Model.....	288
12.2 Estimating the Model Parameters	292
12.3 Assessing the Importance of the Model	296
12.4 Properties of the Model Parameters.....	299
12.5 Using the Fitted Model to Predict Responses.....	301
12.6 Summarizing the Fit of the Model	305
12.7 Consequences of Uncertainty in the Explanatory Variate	306
12.8 Using Replication to Test Goodness of Fit.....	308
12.9 Variations on the Model	313
12.9.1 Centering and Scaling the Explanatory Variate	313
12.9.2 Regression through the Origin.....	314
12.9.3 Calibration	320
Exercises	321

13. Checking Model Fit	325
13.1 Checking the Form of the Model	325
13.2 More Ways of Estimating Deviations	328
13.3 Using Graphical Tools to Check Assumptions	330
13.4 Looking for Influential Observations	332
13.4.1 Measuring Potential Influence: Leverage	333
13.4.2 The Relationship between Residuals and Leverages	335
13.4.3 Measuring the Actual Influence of Individual Observations	336
13.5 Assessing the Predictive Ability of a Model: Cross-Validation	338
Exercises	342
14. Models for Several Variates: Multiple Linear Regression	345
14.1 Visualizing Relationships between Variates	345
14.2 Defining the Model	347
14.3 Estimating the Model Parameters	350
14.4 Assessing the Importance of Individual Explanatory Variates	352
14.4.1 Adding Terms into the Model: Sequential ANOVA and Incremental Sums of Squares	353
14.4.2 The Impact of Removing Model Terms: Marginal Sums of Squares	356
14.5 Properties of the Model Parameters and Predicting Responses	358
14.6 Investigating Model Misspecification	359
14.7 Dealing with Correlation among Explanatory Variates	361
14.8 Summarizing the Fit of the Model	365
14.9 Selecting the Best Model	366
14.9.1 Strategies for Sequential Variable Selection	369
14.9.2 Problems with Procedures for the Selection of Subsets of Variables	376
14.9.3 Using Cross-Validation as a Tool for Model Selection	377
14.9.4 Some Final Remarks on Procedures for Selecting Models	378
Exercises	378
15. Models for Variates and Factors	381
15.1 Incorporating Groups into the Simple Linear Regression Model	382
15.1.1 An Overview of Possible Models	383
15.1.2 Defining and Choosing between the Models	388
15.1.2.1 Single Line Model	388
15.1.2.2 Parallel Lines Model	388
15.1.2.3 Separate Lines Model	390
15.1.2.4 Choosing between the Models: The Sequential ANOVA Table	391
15.1.3 An Alternative Sequence of Models	396
15.1.4 Constraining the Intercepts	398
15.2 Incorporating Groups into the Multiple Linear Regression Model	399
15.3 Regression in Designed Experiments	406
15.4 Analysis of Covariance: A Special Case of Regression with Groups	409
15.5 Complex Models with Factors and Variates	414
15.5.1 Selecting the Predictive Model	414
15.5.2 Evaluating the Response: Predictions from the Fitted Model	417

15.6	The Connection between Factors and Variates	417
15.6.1	Rewriting the Model in Matrix Notation	421
	Exercises	423
16.	Incorporating Structure: Linear Mixed Models	427
16.1	Incorporating Structure	427
16.2	An Introduction to Linear Mixed Models	428
16.3	Selecting the Best Fixed Model	430
16.4	Interpreting the Random Model	432
16.4.1	The Connection between the Linear Mixed Model and Multi-Stratum ANOVA	434
16.5	What about Random Effects?	435
16.6	Predicting Responses	436
16.7	Checking Model Fit	437
16.8	An Example	438
16.9	Some Pitfalls and Dangers	444
16.10	Extending the Model	445
	Exercises	447
17.	Models for Curved Relationships	451
17.1	Fitting Curved Functions by Transformation	451
17.1.1	Simple Transformations of an Explanatory Variate	451
17.1.2	Polynomial Models	456
17.1.3	Trigonometric Models for Periodic Patterns	460
17.2	Curved Surfaces as Functions of Two or More Variates	463
17.3	Fitting Models Including Non-Linear Parameters	472
	Exercises	476
18.	Models for Non-Normal Responses: Generalized Linear Models	479
18.1	Introduction to Generalized Linear Models	480
18.2	Analysis of Proportions Based on Counts: Binomial Responses	481
18.2.1	Understanding and Defining the Model	483
18.2.2	Assessing the Importance of the Model and Individual Terms: The Analysis of Deviance	487
18.2.2.1	Interpreting the ANODEV with No Over-Dispersion	489
18.2.2.2	Interpreting the ANODEV with Over-Dispersion	490
18.2.2.3	The Sequential ANODEV Table	493
18.2.3	Checking the Model Fit and Assumptions	494
18.2.4	Properties of the Model Parameters	496
18.2.5	Evaluating the Response to Explanatory Variables: Prediction	498
18.2.6	Aggregating Binomial Responses	500
18.2.7	The Special Case of Binary Data	501
18.2.8	Other Issues with Binomial Responses	501
18.3	Analysis of Count Data: Poisson Responses	502
18.3.1	Understanding and Defining the Model	503
18.3.2	Analysis of the Model	506
18.3.3	Analysing Poisson Responses with Several Explanatory Variables	509
18.3.4	Other Issues with Poisson Responses	512

18.4 Other Types of GLM and Extensions	512
Exercises	513
19. Practical Design and Data Analysis for Real Studies	517
19.1 Designing Real Studies	518
19.1.1 Aims, Objectives and Choice of Explanatory Structure	518
19.1.2 Resources, Experimental Units and Constraints	519
19.1.3 Matching the Treatments to the Resources	520
19.1.4 Designs for Series of Studies and for Studies with Multiple Phases	521
19.1.5 Design Case Studies	523
19.2 Choosing the Best Analysis Approach	535
19.2.1 Analysis of Designed Experiments	536
19.2.2 Analysis of Observational Studies	537
19.2.3 Different Types of Data	538
19.3 Presentation of Statistics in Reports, Theses and Papers	538
19.3.1 Statistical Information in the Materials and Methods	539
19.3.2 Presentation of Results	540
19.4 And Finally.	543
References	545
Appendix A: Data Tables	551
Appendix B: Quantiles of Statistical Distributions	559
Appendix C: Statistical and Mathematical Results	563
Index	569