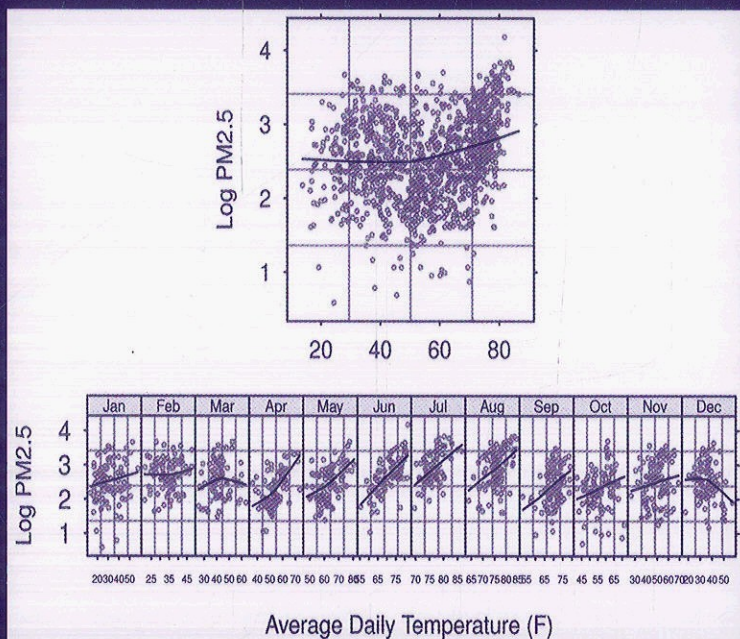


CHAPMAN & HALL/CRC
APPLIED ENVIRONMENTAL STATISTICS

ENVIRONMENTAL AND ECOLOGICAL STATISTICS WITH R



SONG S. QIAN



CRC Press
Taylor & Francis Group

A CHAPMAN & HALL BOOK

Contents

Preface	v
Table of Contents	ix
List of Tables	xiii
List of Figures	xv
I Basic Concepts	1
1 Introduction	3
1.1 The Everglades Example	6
1.2 Statistical Issues	9
1.3 Bibliography Notes	12
2 R	13
2.1 What is R?	13
2.2 Getting Started with R	13
2.2.1 R Prompt and Assignment	14
2.2.2 Data Types	15
2.2.3 R Functions	16
2.3 The R Commander	18
3 Statistical Assumptions	25
3.1 The Normality Assumption	25
3.2 The Independence Assumption	29
3.3 The Constant Variance Assumption	30
3.4 Exploratory Data Analysis	32
3.4.1 Graphs for Displaying Distributions	32
3.4.2 Graphs for Comparing Distributions	35
3.4.3 Graphs for Exploring Dependency Among Variables	36
3.5 From Graphs to Statistical Thinking	45
3.6 Bibliography Notes	47
4 Statistical Inference	49
4.1 Estimation of Population Mean and Confidence Interval	50
4.1.1 Bootstrap Method for Estimating Standard Error	57
4.2 Hypothesis Testing	61

4.2.1	<i>T</i> -Test	62
4.2.2	Two-Sided Alternatives	69
4.2.3	Hypothesis Testing Using the Confidence Interval	70
4.3	A General Procedure	72
4.4	Nonparametric Methods for Hypothesis Testing	73
4.4.1	Rank Transformation	73
4.4.2	Wilcoxon Signed Rank Test	74
4.4.3	Wilcoxon Rank Sum Test	75
4.4.4	A Comment on Distribution-Free Methods	77
4.5	Significance Level α , Power $1 - \beta$, and <i>p</i> -Value	80
4.6	One-Way Analysis of Variance	87
4.6.1	Analysis of Variance	88
4.6.2	Statistical Inference	90
4.6.3	Multiple Comparisons	92
4.7	Examples	98
4.7.1	The Everglades Example	98
4.7.2	Kemp's Ridley Turtles	99
4.7.3	Assessing Water Quality Standard Compliance	105
4.7.4	Interaction between Red Mangrove and Sponges	108
4.8	Bibliography Notes	113

II Statistical Modeling 115

5 Linear Models 119

5.1	ANOVA as a Linear Model	122
5.2	Simple and Multiple Linear Regression Models	124
5.2.1	The Least Squares	125
5.2.2	PCBs in the Fish Example	126
5.2.3	Regression with One Predictor	127
5.2.4	Multiple Regression	129
5.2.5	Interaction	131
5.2.6	Residuals and Model Assessment	133
5.2.7	Categorical Predictors	140
5.2.8	The Finnish Lakes Example and Collinearity	144
5.3	General Considerations in Building a Predictive Model	155
5.4	Uncertainty in Model Predictions	159
5.5	Two-Way ANOVA	161
5.5.1	Interaction	166
5.6	Bibliography Notes	167

6 Nonlinear Models 169

6.1	Nonlinear Regression	169
6.1.1	Piecewise Linear Models	178
6.1.2	Example: U.S. Lilac First Bloom Dates	184
6.2	Smoothing	188

6.2.1	Scatter Plot Smoothing	188
6.2.2	Fitting a Local Regression Model	191
6.3	Smoothing and Additive Models	193
6.3.1	Additive Models	193
6.3.2	Fitting an Additive Model	196
6.3.3	The North American Wetland Database	198
6.3.4	Discussion: The Role of Nonparametric Regression Models in Science	199
6.3.5	Seasonal Decomposition of Time Series	206
6.4	Bibliographic Notes	215
7	Classification and Regression Tree	217
7.1	The Willamette River Example	217
7.2	Statistical Methods	221
7.2.1	Growing and Pruning a Regression Tree	223
7.2.2	Growing and Pruning a Classification Tree	232
7.2.3	Plotting Options	234
7.3	Comments	240
7.3.1	CART as a Model Building Tool	240
7.3.2	Deviance and Probabilistic Assumptions	244
7.3.3	CART and Ecological Threshold	245
7.4	Bibliography Notes	247
8	Generalized Linear Model	249
8.1	Logistic Regression	250
8.1.1	Example: Evaluating the Effectiveness of UV as a Drinking Water Disinfectant	251
8.1.2	Statistical Issues	252
8.1.3	Fitting the Model in R	253
8.2	Model Interpretation	255
8.2.1	Logit Transformation	255
8.2.2	Intercept	256
8.2.3	Slope	256
8.2.4	Additional Predictors	257
8.2.5	Interaction	259
8.2.6	Comments on the Crypto Example	260
8.3	Diagnostics	261
8.3.1	Binned Residuals Plot	261
8.3.2	Overdispersion	262
8.4	Seed Predation by Rodents: A Second Example of Logistic Regression	264
8.5	Poisson Regression Model	277
8.5.1	Arsenic Data from Southwestern Taiwan	277
8.5.2	Poisson Regression	278
8.5.3	Exposure and Offset	285

8.5.4	Overdispersion	286
8.5.5	Interactions	289
8.5.6	Poisson Regression versus Logistic Regression	295
8.5.7	Negative Binomial	299
8.6	Generalized Additive Models	301
8.6.1	Example: Whales in the Western Antarctic Peninsula	303
8.7	Bibliography Notes	314
III	Advanced Statistical Modeling	315
9	Simulation for Model Checking and Statistical Inference	317
9.1	Simulation	317
9.2	Summarizing Linear and Nonlinear Regression Using Simulation	319
9.2.1	An Introductory Example	319
9.2.2	Summarizing a Linear Regression Model	322
9.2.3	Simulation for Model Evaluation	326
9.3	Simulation Based on Re-sampling	333
9.3.1	Bootstrap Aggregation	334
9.3.2	Example: Confidence Interval of the CART-Based Threshold	335
9.4	Bibliography Notes	338
10	Multilevel Regression	339
10.1	Multilevel Structure and Exchangeability	340
10.2	Multilevel ANOVA	343
10.2.1	Intertidal Seaweed Grazers	344
10.2.2	Background N ₂ O Emission from Agriculture Fields	348
10.2.3	When to Use the Multilevel Model?	352
10.2.4	Two-Way ANOVA	354
10.3	Multilevel Linear Regression	362
10.3.1	Nonnested Groups	373
10.3.2	Multiple Regression Problems	378
10.4	Generalized Multilevel Models	390
10.4.1	Liverpool Moths – A Logistic Regression Example	390
10.4.2	Cryptosporidium in U.S. Drinking Water – A Poisson Regression Example	395
10.4.3	Model Checking Using Simulation	400
10.5	Bibliography Notes	407
	Bibliography	408
	Index	417