

SURVEYING NATURAL POPULATIONS

Quantitative Tools for Assessing Biodiversity

SECOND EDITION



LEE-ANN C. HAYEK & MARTIN A. BUZAS

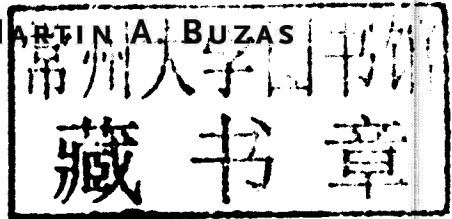
Surveying

Natural Populations

QUANTITATIVE TOOLS FOR ASSESSING BIODIVERSITY

SECOND EDITION

LEE-ANN C. HAYEK & MARTIN A. BUZAS



COLUMBIA UNIVERSITY PRESS
NEW YORK

COLUMBIA UNIVERSITY PRESS
PUBLISHERS SINCE 1893
NEW YORK CHICHESTER, WEST SUSSEX
COPYRIGHT © 2010 THE SMITHSONIAN INSTITUTION
ALL RIGHTS RESERVED

LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA
HAYEK, LEE-ANN C.
SURVEYING NATURAL POPULATIONS : QUANTITATIVE TOOLS FOR ASSESSING
BIODIVERSITY / LEE-ANN C. HAYEK AND MARTIN A. BUZAS. — 2ND ED.
P. CM.
ISBN 978-0-231-14620-3 (CLOTH : ALK. PAPER)
1. POPULATION BIOLOGY—STATISTICAL METHODS. 2. PALEOECOLOGY—
STATISTICAL METHODS. I. BUZAS, MARTIN A. II. TITLE.

QH532.H39 2010
577.8'80727—DC22

2010009475



COLUMBIA UNIVERSITY PRESS BOOKS ARE PRINTED ON PERMANENT AND
DURABLE ACID-FREE PAPER.
THIS BOOK IS PRINTED ON PAPER WITH RECYCLED CONTENT.
PRINTED IN THE UNITED STATES OF AMERICA

c 10 9 8 7 6 5 4 3 2 1

REFERENCES TO INTERNET WEB SITES (URLS) WERE ACCURATE AT THE TIME OF WRITING. NEITHER
THE AUTHOR NOR COLUMBIA UNIVERSITY PRESS IS RESPONSIBLE FOR URLS THAT MAY HAVE
EXPIRED OR CHANGED SINCE THE MANUSCRIPT WAS PREPARED.

To Christine, Matthew, Katherine, Aron, and my new daughter Kate

and

To Barbara, Pamela, Jeffrey, and Thomas

PREFACE

WE BEGAN THIS WORK as a revision of our text *Surveying Natural Populations* published in 1997. Fortunately, our colleagues at many universities and colleges kindly identified textual errors or misleading writing that we incorporated. Those who used the book as a classroom text and/or reference gave especially helpful advice. However, as we began this updating and especially when writing about the exciting new and newly developed material on SHE analysis, we realized that this book would require not just a major revision, but also new chapters. In particular, we incorporated entire developmental chapters on quantitative biodiversity. This book now contains 7 chapters on the analysis and interpretation of the quantitative results that are obtained from a study of biodiversity. Chapters 12 through 18 provide a synthesis of the entire field of biodiversity assessment. Most importantly, in this book all of the major ecological and biodiversity indices have been related mathematically under the umbrella of information measures. We show how the use of a single measure is not a solitary numerical calculation but a purposeful selection of a member of a mathematical family, each selection having repercussions for the interpretation of the resultant biodiversity data. We also demonstrate how many suggested diversity indices must be rejected because they lack vital properties and they do not follow an information-theoretic scheme wherein all measures are mathematically linked. Use of such a scheme not only relates all measures within a study but also provides the standardization necessary for individual studies to be related over time or space allowing for more insight and stronger generalizations from biodiversity analyses.

In the intervening decade since the first book was published, the plight of both humans and other species has been brought more clearly to the attention of both scientists and the public. Our declining global conditions of anthropogenic origin are sending imperatives to those who study natural populations; we can no longer utilize qualitative assessments of biodiversity. We must use the latest and most mathematically useful methods for the evaluation of affected and potentially affected species assemblages and related environmental conditions. We believe that this book is needed now more than ever.

As we stated in the Preface to our original book, the degradation of environments and the fate of natural populations that inhabit them has become a topic of paramount concern throughout the world. Obtaining collections from the field along with pertinent observations, as well as the identification and enumeration of organisms, falls

within the domain of field biologists. Because of the enormous quantities involved, all organisms cannot be directly counted and population parameters must be estimated. Estimation from sample to population falls within the domain of mathematical statisticians. Unfortunately, the two groups of researchers often do not communicate well. This book attempts to bridge the gap.

One author (LCH) is a mathematical statistician with extensive experience in statistical theory and practice for developing and interpreting statistical science for solving ecological, natural history, and biodiversity problems; the other (MAB) is a researcher in quantitative ecology–paleoecology who is familiar with statistics usage. Their knowledge and experience form overlapping sets. Together, they integrate the intuition of the experienced field researcher with the statistical rationale required to survey natural populations.

This is not a book for statisticians. The authors do not offer mathematical proofs of their results. Statistical knowledge is extrapolated without the type of documentation that mathematical statisticians expect. The concepts and complexities contained in the area of mathematical application are, however, requisite knowledge for field biologists and conservation biologists with marine or terrestrial specialties as well as for site managers, paleontologists, and archaeologists. This book is for you.

No prior statistical knowledge is assumed, and only a familiarity with, or, at least not an aversion to algebra is required. This is a versatile book that can be used at many levels: (1) serious students can learn biological and statistical principles fundamental to sound quantitative surveys; (2) experienced researchers can, for the first time in a single readable source, find the underlying statistical theory for their commonly used field approaches; and (3) graduate students and researchers can discover many new relationships, derivations, and measures that can open new research areas suggested throughout the book. For field researchers in natural systems, this book is a reference, a text, a tool, and a guide.

ACKNOWLEDGMENTS

WE THANK F. DALLMEIER, Director of the Smithsonian Man and the Biosphere Program at the Smithsonian's National Zoological Park, for graciously providing the primary data sets used in this book. Data sets and information on the Indian River field project were provided by S. Reed of the Smithsonian's Marine Station at Ft. Pierce, Florida, and for the amphibian work by R. Heyer of the Smithsonian's National Museum of Natural History.

Without J. Jett this book could never have been completed; we thank her for her invaluable assistance, organization, checking, for her exceptional effort of compiling the index and for putting up with us and our hectic schedule. A. Hayek entered data for some of the problems and ran some analyses for the solutions. We thank M. Parrish from the NMNH Department of Paleobiology for the cover and interior artwork.

SURVEYING NATURAL POPULATIONS

CONTENTS

Preface xix

Acknowledgments xxi

1. INTRODUCTION 1

KEY CONCEPTS 2

HOW TO USE THIS BOOK 6

The Illustrative Example 6

An Overview of Terminology and Statistical Methods 7

Mathematical Equations and End-of-Chapter Problems 9

2. DENSITY: MEAN AND VARIANCE 11

MEAN DENSITY 12

RANDOM SELECTION 14

VARIANCE AND STANDARD DEVIATION
FOR DENSITY ESTIMATES 16

INTERPRETING VARIANCE OF SMALL AND LARGE SAMPLES:
TCHEBYCHEV'S THEOREM 19

SUMMARY 20

PROBLEMS 20

3. NORMAL AND SAMPLING DISTRIBUTIONS FOR FIELDWORK 22

FIELD DATA SUMMARY 22

Histograms 23

The Normal Distribution 25

Standardized Normal Distribution 26

SAMPLING DISTRIBUTION OF THE MEAN 29

Combinatorial Analysis 30

Mean of the Sampling Distribution 31

THE STANDARD ERROR 32

THE CENTRAL LIMIT THEOREM 35

SUMMARY 36

PROBLEMS 37

4. CONFIDENCE LIMITS AND INTERVALS FOR DENSITY 38

CONFIDENCE LIMITS ON MEAN DENSITY 38

Limits and Intervals for Large Samples 39

Limits and Intervals for Small Samples 41

Limits and Intervals for Total Population 45

PRECISION OF THE CONFIDENCE INTERVAL 46

INTERPRETATION OF CONFIDENCE LIMITS
AND INTERVALS 49

SUMMARY 50

PROBLEMS 51

5. HOW MANY FIELD SAMPLES? 53

MINIMALLY SIZED SAMPLES WITH NO PILOT DATA 53

Law of Large Numbers 53

Sample Size with No Assumption of Normality 54

Sample Size with Assumption of Normality 55

MINIMALLY SIZED SAMPLES WITH PRELIMINARY
PILOT DATA 57

Sample Size with an Estimate of the Standard Deviation 57

Large Versus Small Sample Results 59

Sample Size with a Bound on Mean Density 60

QUADRAT SIZE OR AREA 61

Homogeneous Case: Bolivian Trees 61

General Cases 63

SUMMARY 67

PROBLEMS 67

6. SPATIAL DISTRIBUTION: THE POWER CURVE 69

SPATIAL RANDOMNESS 70

SPATIAL NONRANDOMNESS 72

POWER CURVE 73

LEAST SQUARES CRITERION FOR THE POWER CURVE COMPUTATION 74

POWER CURVE FOR ESTIMATING SPATIAL DISTRIBUTION 75

Minimizing the Error for All Species 78

Logarithmic Transformation 79

Spatial Distribution of Individuals 79

Comparison of Differently Sized Samples of Individuals:

Scale Changes 80

Combining Samples of Different Sizes: Problems of Scale 81

EXAMPLES OF POWER CURVE CALCULATIONS FOR ORGANISMS 82

THE SLOPE AS AN INDEX OF AGGREGATION 84

THE IMPORTANCE OF THE INTERCEPT a 84

USING THE POWER CURVE TO CHOOSE AN OPTIMAL SAMPLE SIZE 87

CHOOSING QUADRAT SIZE WHEN POWER CURVE PARAMETERS
ARE UNKNOWN 89

SUMMARY 90

PROBLEMS 91

7. FIELD SAMPLING SCHEMES 93

SIMPLE (UNRESTRICTED) RANDOM SAMPLING 93

THE QUADRAT AS THE PRIMARY SAMPLING UNIT 95

THE INDIVIDUAL AS THE PRIMARY SAMPLING UNIT 95

CLUSTER SAMPLING 96

SYSTEMATIC (OR ORDERED) SAMPLING 97

Estimators for Systematic Sampling 98

Sampling Distribution for Systematic Sample Collection 100

The Mean of a Systematic Sample 101

The Variance of a Systematic Sample 102

REPLICATED SYSTEMATIC SAMPLING 103

TARGET POPULATION, SPATIAL DISTRIBUTION, AND
SYSTEMATIC SAMPLING 105

SAMPLE SELECTION STRATEGIES: RANDOM VERSUS SYSTEMATIC	105
STRATIFIED SAMPLING	107
COMPARISON OF SYSTEMATIC AND STRATIFIED SAMPLING	109
Systematic Samples as Implicit Strata	109
Systematic Sampling for Stratum Detection	109
Use of Supplemental Species Data from the Target Population	109
Sampling Within Each Stratum	110
STRATIFICATION AND THE SIZE OF THE TARGET POPULATION	111
EXAMPLE OF STRATIFICATION	113
CAUTIONS FOR STRATIFICATION	116
DESIGN INTERRELATIONSHIPS	117
CHOOSING THE BEST SAMPLING DESIGN	120
SUMMARY	122

8. SPECIES PROPORTIONS: RELATIVE ABUNDANCES 124

PROBABILITY OF A RANDOMLY SELECTED INDIVIDUAL BELONGING TO A SPECIES	124
BERNOULLI TRIALS	126
THE BINOMIAL DISTRIBUTION	126
The Binomial Distribution of Counts	127
The Binomial Distribution of Proportions	134
Examining the Parameters of the Binomial Distribution	137
THE POISSON APPROXIMATION	139
THE NORMAL APPROXIMATION	140
CONFIDENCE LIMITS FOR SPECIES PROPORTIONS	142
MINDING P's AND Q's	143
AN EXAMPLE OF SPECIES PROPORTION: ESTIMATION FROM THE BOLIVIAN TREES	145
A BIOLOGICALLY REPRESENTATIVE SAMPLE SIZE	147
STATISTICAL SAMPLE SIZE FOR RELATIVE ABUNDANCES OF SPECIES	149
CLUSTER SAMPLING FOR SPECIES PROPORTIONS	151
A FOSSIL EXAMPLE OF CONFIDENCE INTERVAL ESTIMATION FOR RELATIVE ABUNDANCE	155

A NOTE OF CAUTION 160

SUMMARY 161

PROBLEMS 162

9. SPECIES DISTRIBUTIONS 164

COVARIANCE 164

MATRIX NOTATION 166

HISTORICAL PERSPECTIVE 171

SERIES AND SEQUENCES 172

HARMONIC SERIES 177

GEOMETRIC SERIES 177

NEGATIVE BINOMIAL DISTRIBUTION 179

 Description of Over-Dispersion 179

 Relationship to the Binomial Distribution 180

 Parameter Estimates 181

 Problems of Fit: A Bolivian Example 186

CONVERGENCE OF NEGATIVE BINOMIAL AND OTHER DISTRIBUTIONS 190

BROKEN STICK DISTRIBUTION 191

LOGARITHMIC SERIES 192

 An Illustration of Calculating α for the Bolivian Data 195

 Class Interval Construction 196

LOG NORMAL DISTRIBUTION 200

 The Need for a Logarithmic Transformation 200

 A Problem of Terminology 201

 Preston's Fit of a Log Normal Distribution 201

 Preston's Rationale and Interval Grouping 202

 Log Normal Fit to Bolivian Data 203

 Density Function Fit 206

LOG TRANSFORMATION PROBLEMS 207

THE CANONICAL LOG NORMAL 209

ESTIMATING TOTAL SPECIES FROM LOG NORMAL 210

CRITIQUE OF THE VARIOUS ABUNDANCE MODELS 211

SUMMARY 213

PROBLEMS 215

10. REGRESSION: OCCURRENCES AND DENSITY 216

RELATIONSHIP OF DENSITY AND OCCURRENCES	216
DESCRIPTIVE STATISTICS OF REGRESSION	217
PROPERTIES OF OCCURRENCES AS A PREDICTOR VARIABLE	218
OBSERVED TREND	220
REGRESSION AND CORRELATION	221
Evaluation of Prediction Error	221
Correlation Coefficient for Regression	224
Interpretation of Predictive Error Measure	225
Coefficient of Determination	226
INFERENCEAL USES OF REGRESSION	227
PREDICTION WITH MEASUREMENT ERROR IN THE OCCURRENCE DATA	229
RELATIONSHIP OF DENSITY AND OCCURRENCES OVER THE TARGET AREA	230
PREDICTION OF DENSITY FROM A SAMPLE OF LOCALITY OCCURRENCES	231
The Bolivian Example: Terrestrial, Small Spatial Scale	231
A Gulf of Mexico Example: Marine, Large Spatial Scale	232
SUMMARY	238
PROBLEMS	239

11. SPECIES OCCURRENCES 240

SURROGATE RANK ORDER OF SPECIES	241
USES OF SMALL SPECIES LISTS OR SMALL SAMPLES OF OCCURRENCES	242
CONFIDENCE LIMITS ON OCCURRENCES	245
SAMPLE SIZE FOR OCCURRENCES	246
OCCURRENCES AS SURROGATE FOR ABUNDANCE DISTRIBUTION	247
FITTING SPECIES ABUNDANCE DISTRIBUTIONS TO OCCURRENCES	248
Sampling the Log Series: An Example	251
Using the Log Series for Occurrence Data: A Cretaceous Example	252
SUMMARY	252
PROBLEMS	254

12. SPECIES DIVERSITY: THE NUMBER OF SPECIES 255

MEAN NUMBER OF SPECIES 256

NUMBER OF DIFFERENT SPECIES 257

SPECIES EFFORT CURVE 258

STANDARDIZING THE NUMBER OF INDIVIDUALS IN A SAMPLE 260

Standardizing in the Data Collection Stage: Standard Counts 260

Standardizing During the Analysis Stage: By Inference 262

REGRESSION OF N AND S 262REGRESSION OF AREA AND S : SPECIES–AREA REGRESSION 264

RAREFACTION 266

Sample Size Standardization by Regression Methods 266

Sanders's Method of Rarefaction 266

The Hypergeometric Distribution for Rarefaction 268

The Binomial Distribution for Rarefaction 271

Equivalent Alpha Diversity for Rarefaction 272

Comparison of Rarefaction Methods 273

ABUNDIFACTION 273

 Regression and Power Curves for Extending
 Species–Area Predictions 274

Extending Species Predictions with a Ratio of Log Relationships 278

Equivalent Alpha Diversity for Abundifaction 280

SUMMARY 282

PROBLEMS 283

13. BIODIVERSITY: DIVERSITY INDICES USING N AND S 284MARGALEF'S d 285LOG–LOG INDEX, $IOTA$ 288

FISHER'S ALPHA 290

FISHER'S ALPHA AND THE LOG SERIES DISTRIBUTION 291

VARIANCE AND STANDARD ERROR OF FISHER'S α INDEX 293EVALUATION OF BIODIVERSITY MEASURES ON OBSERVATIONS
 WITH N AND S ONLY 295

SUMMARY 295

PROBLEM 296

14. BIODIVERSITY: DIVERSITY MEASURES USING RELATIVE ABUNDANCES 297

SPECIES ABUNDANCES	297
THE RELATIVE SPECIES ABUNDANCE VECTOR, RSAV, AND NOTATION	298
SIMPLIFYING SUMMARY NOTATION	299
PLOTTING THE DATA	300
RICHNESS MEASURES AND THE RSAV	301
EFFICACY AND RULES FOR DIVERSITY MEASURES	301
DIVERSITY MEASURES WITH NO ASSUMPTION ABOUT STATISTICAL DISTRIBUTION	303
SIMPSON'S LAMBDA	303
BERGER AND PARKER INDEX	306
INFORMATION	307
INFORMATION FROM NATURAL POPULATIONS	308
SHANNON'S INFORMATION FUNCTION	309
H AS A DIVERSITY MEASURE	310
GENERALIZATION OF DIVERSITY MEASUREMENT	315
SUMMARY	318
PROBLEMS	319

15. BIODIVERSITY: DOMINANCE AND EVENNESS 320

DEVELOPMENT OF GENERALIZED NOTATION FOR EVENNESS AND DOMINANCE	321
EVENNESS MEASURES	324
EVENNESS MEASURES RELATED AS SPECIAL CASES IN A SINGLE FAMILY	324
AN ALTERNATIVE FAMILY OF EVENNESS MEASURES	326
COMPARING EVENNESS VALUES	327
The Repeat, or Mixtures of Distributions	328
Changes in Simpson's and Shannon's Measures Under Repeat	329
The Extremes of Evenness Indices	331
E' and the Repeat, or Mixtures	334

SUMMARY 334

PROBLEMS 335

16. BIODIVERSITY: UNIFYING DIVERSITY AND EVENNESS MEASURES WITH CANONICAL EQUATIONS 337

HISTORICAL DERIVATION OF DECOMPOSITION 337

DECOMPOSITION AND THE LOG CONTINUUM 338

CHECKING THE UNIFIED MEASURES WITH A NUMERICAL EXAMPLE 339

DECOMPOSITION WITH E' MEASURES 340

SUMMARY 340

PROBLEMS 341

17. BIODIVERSITY: SHE ANALYSIS AS THE ULTIMATE UNIFICATION THEORY OF BIODIVERSITY WITH THE COMPLETE BIODIVERSITYGRAM 342

SHE ANALYSIS—AN OVERVIEW 343

SHE AS DISTRIBUTION-FREE METHODOLOGY 343

ACCUMULATING OBSERVED SAMPLES 344

THE BIODIVERSITYGRAM: BDG 344

FAMILY OF INFORMATION-BASED MEASURES AND ENTROPY 346

ALTERNATIVE CONCEPTUALIZATIONS OF EVENNESS 347

 Evenness as a Residual 347

 Evenness as Redundancy 348

 Evenness as Uncertainty 349

A PERCENTAGE MEASURE OF DISTANCE FROM UNIFORM ABUNDANCES 349

SHE AS COMPLETE INFERENTIAL METHOD FOR BIODIVERSITY ANALYSIS AND EVALUATION 349

FITTING STATISTICAL DISTRIBUTIONS: GOODNESS OF FIT WITH FREQUENCY VERSUS DISTRIBUTION FUNCTION 350

 Frequency Function Fit 350

 Distribution Function Fit 351

SUMMARY 352

PROBLEMS 353