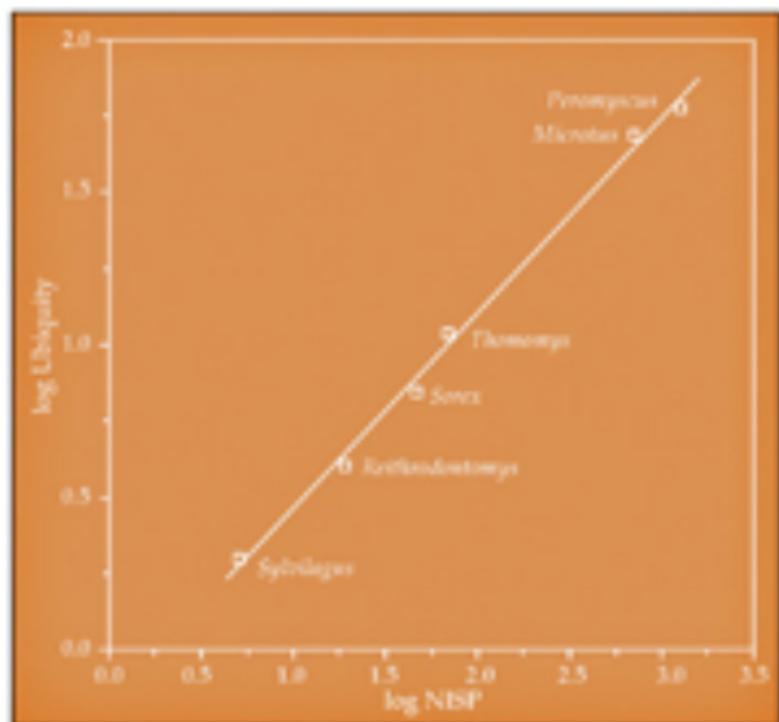


# Quantitative Paleozoology

R. Lee Lyman



This page intentionally left blank

## Quantitative Paleozoology

*Quantitative Paleozoology* describes and illustrates how the remains of long-dead animals recovered from archaeological and paleontological excavations can be studied and analyzed. The methods range from determining how many animals of each species are represented to determining whether one collection consists of more broken and more burned bones than another. All methods are described and illustrated with data from real collections, while numerous graphs illustrate various quantitative properties.

R. LEE LYMAN is professor of anthropology at the University of Missouri-Columbia. A scholar of late Quaternary paleomammalogy and human prehistory of the Pacific Northwest United States, he is the author of *Vertebrate Taphonomy*, and, most recently, the coeditor of *Zooarchaeology and Conservation Biology*.



**Cambridge Manuals in Archaeology**

*General Editor*

Graeme Barker, *University of Cambridge*

*Advisory Editors*

Elizabeth Slater, *University of Liverpool*

Peter Bogucki, *Princeton University*

*Cambridge Manuals in Archaeology* is a series of reference handbooks designed for an international audience of upper-level undergraduate and graduate students and for professional archaeologists and archaeological scientists in universities, museums, research laboratories, and field units. Each book includes a survey of current archaeological practice alongside essential reference material on contemporary techniques and methodology.

*Books in the series*

*Pottery in Archaeology*, CLIVE ORTON, PAUL TYERS, and ALAN VINCE

*Vertebrate Taphonomy*, R. LEE LYMAN

*Photography in Archaeology and Conservation*, 2nd edition, PETER G. DORRELL

*Alluvial Geoarchaeology*, A. G. BROWN

*Shells*, CHERYL CLAASEN

*Sampling in Archaeology*, CLIVE ORTON

*Excavation*, STEVE ROSKAMS

*Teeth*, 2nd edition, SIMON HILLSON

*Lithics*, 2nd edition, WILLIAM ANDREFSKY, JR.

*Geographical Information Systems in Archaeology*, JAMES CONOLLY and MARK LAKE

*Demography in Archaeology*, ANDREW CHAMBERLAIN

*Analytical Chemistry in Archaeology*, A. M. POLLARD, C. M. BATT, B. STERN, and S. M. M. YOUNG

*Zooarchaeology*, 2nd edition, ELIZABETH J. REITZ and ELIZABETH S. WING



# Quantitative Paleozoology

R. Lee Lyman *University of Missouri-Columbia*



CAMBRIDGE UNIVERSITY PRESS  
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press  
The Edinburgh Building, Cambridge CB2 8RU, UK  
Published in the United States of America by Cambridge University Press, New York  
[www.cambridge.org](http://www.cambridge.org)  
Information on this title: [www.cambridge.org/9780521887496](http://www.cambridge.org/9780521887496)

© R. Lee Lyman 2008

This publication is in copyright. Subject to statutory exception and to the provision of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published in print format 2008

ISBN-13 978-0-511-38646-6 eBook (EBL)

ISBN-13 978-0-521-88749-6 hardback

ISBN-13 978-0-521-71536-2 paperback

Cambridge University Press has no responsibility for the persistence or accuracy of urls for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

# CONTENTS

---

<i>List of figures</i>	<i>page</i> xi
<i>List of tables</i>	xvii
<i>Preface</i>	xxi
<b>1. Tallying and Counting: Fundamentals</b>	1
Paleozoological Concepts	4
Mathematical and Statistical Concepts	8
Scales of Measurement	8
Measured and Target Variables: Reliability and Validity	11
Absolute and Relative Frequencies and Closed Arrays	13
Discussion	16
Background of Some Faunal Samples	17
<b>2. Estimating Taxonomic Abundances: NISP and MNI</b>	21
The Number of Identified Specimens (NISP)	27
Advantages of NISP	28
Problems with NISP	29
Problems, Schmblems	30
A Problem We <i>Should</i> Worry About	36
The Minimum Number of Individuals (MNI)	38
Strengths(?) of MNI	43
Problems with MNI	45
Aggregation	57
Defining Aggregates	67
Discussion	69
Which Scale of Measurement?	71
Resolution	78
Conclusion	81

<b>3. Estimating Taxonomic Abundances: Other Methods</b>	83
Biomass and Meat Weight	84
Measuring Biomass	85
Problems with Measuring Biomass (based on MNI)	86
Solving Some Problems in Biomass Measurement	88
Measuring Meat Weight	89
The Weight Method (Skeletal Mass Allometry)	93
Bone Weight	102
Bone Size and Animal Size Allometry	108
Ubiquity	114
Matching and Pairing	119
More Pairs Means Fewer Individuals	121
The Lincoln–Petersen Index	123
Identifying Bilateral Pairs	129
Correcting for Various Things	134
Size	137
Discussion	139
<b>4. Sampling, Recovery, and Sample Size</b>	141
Sampling to Redundancy	143
Excavation Amount	144
NISP as a Measure of Sample Redundancy	146
Volume Excavated or NISP	149
The Influences of Recovery Techniques	152
Hand Picking Specimens by Eye	152
Screen Mesh Size	154
To Correct or Not to Correct for Differential Loss	156
Summary	158
The Species–Area Relationship	159
Species–Area Curves Are Not All the Same	164
Nestedness	167
Conclusion	171
<b>5. Measuring the Taxonomic Structure and Composition (“Diversity”)</b>	
<b>of Faunas</b>	172
Basic Variables of Structure and Composition	174
Indices of Structure and Similarity	178
Taxonomic Richness	179
Taxonomic Composition	185

Taxonomic Heterogeneity	192
Taxonomic Evenness	194
Discussion	198
Trends in Taxonomic Abundances	203
Conclusion	209
<b>6. Skeletal Completeness, Frequencies of Skeletal Parts, and Fragmentation</b>	214
History of the MNE Quantitative Unit	215
Determination of MNE Values	218
MNE Is Ordinal Scale at Best	222
A Digression on Frequencies of Left and Right Elements	229
Using MNE Values to Measure Skeletal-Part Frequencies	232
Modeling and Adjusting Skeletal-Part Frequencies	233
Measuring Skeletal Completeness	241
A Suggestion	244
Measuring Fragmentation	250
Fragmentation Intensity and Extent	250
The NISP:MNE Ratio	251
Discussion	254
Conclusion	261
<b>7. Tallying for Taphonomy: Weathering, Burning, Corrosion, and Butchering</b>	264
Yet Another Quantitative Unit	266
Weathering	267
Chemical Corrosion and Mechanical Abrasion	273
Burning and Charring	274
A Digression	276
Gnawing Damage	277
Butchering Marks	279
Types of Butchering Damage	280
Tallying Butchering Evidence: General Comments	281
Tallying Percussion Damage	283
Tallying Cut Marks and Cut Marked Specimens	284
The Surface Area Solution	286
Discussion	291
Conclusion	296

<b>8. Final Thoughts</b>	299
Counting as Exploration	302
<i>Glossary</i>	309
<i>References</i>	313
<i>Index</i>	345

# LIST OF FIGURES

---

1.1.	Chester Stock's pie diagram of abundances of five mammalian orders represented in faunal remains from Rancho La Brea.	page 2
2.1.	Schematic illustration of loss and addition to a set of faunal remains studied by a paleozoologist.	23
2.2.	Taxonomic relative abundances across five strata.	33
2.3.	The theoretical limits of the relationship between NISP and MNI.	49
2.4.	The theoretically expected relationship between NISP and MNI.	50
2.5.	Relationship between NISP and MNI data pairs for mammal remains from the Meier site.	52
2.6.	Relationship between NISP and MNI data pairs for the precontact mammal remains from the Cathlapotle site.	53
2.7.	Relationship between NISP and MNI data pairs for the postcontact mammal remains from the Cathlapotle site.	54
2.8.	Relationship between NISP and MNI data pairs for remains of six mammalian genera in eighty-four owl pellets.	56
2.9.	Amount by which a taxon's MNI increases if the minimum distinction MNI is changed to the maximum distinction MNI in eleven assemblages.	60
2.10.	Change in the ratio of deer to gopher abundances in eleven assemblages when MNImax is used instead of MNImin.	61
2.11.	Relationships between NISP and MNImin, and NISP and MNImax at site 45DO214.	66
2.12.	Ratios of abundances of four least common taxa in a collection of eighty-four owl pellets based on NISP, MNImax, and MNImin.	72
2.13.	Frequency distributions of NISP and MNI taxonomic abundances in the Cathlapotle fauna.	73

2.14.	Frequency distributions of NISP and MNI taxonomic abundances in the 45OK258 fauna in eastern Washington State.	74
2.15.	Frequency distributions of NISP and MNI taxonomic abundances in two lumped late-prehistoric mammal assemblages from the western Canadian Arctic.	75
2.16.	Frequency distributions of NISP and MNI taxonomic abundances in four lumped historic era mammalian faunas.	76
3.1.	Ontogenetic, seasonal, and sexual variation in live weight of one male and one female Columbian black-tailed deer.	88
3.2.	Relationship between bone weight per individual and soft-tissue weight in domestic pig.	98
3.3.	Relationship between bone weight per skeletal portion and gross weight per skeletal portion in 6-month-old domestic sheep and 90-month-old domestic sheep.	101
3.4.	Frequency distributions of biomass per taxon in two sites in Florida State.	106
3.5.	Frequency distributions of biomass per mammalian taxon in a site in Georgia State.	107
3.6.	Relationship between lateral length of white-tailed deer astragali and body weight.	112
3.7.	Relationship between NISP and ubiquity of six genera in a collection of eighty-four owl pellets.	115
3.8.	Relationship between NISP and ubiquity of twenty-eight taxa in eighteen sites.	117
3.9.	Relationship between NISP and ubiquity of fifteen taxa in seven analytical units in site 45DO189.	119
3.10.	Relationship between NISP and ubiquity of eighteen taxa in four analytical units in site 45OK2.	120
3.11.	A model of how the Lincoln-Petersen index is calculated.	125
3.12.	Latero-medial width of the distal condyle and minimum antero-posterior diameter of the middle groove of the distal condyle of forty-eight pairs of left and right distal humeri of <i>Odocoileus virginianus</i> and <i>Odocoileus hemionus</i> .	131
3.13.	A model of how two dimensions of a bone can be used to determine degree of (a)symmetry between bilaterally paired left and right elements.	132
4.1.	Cumulative richness of mammalian genera across cumulative volume of sediment excavated annually at the Meier site.	146

4.2.	Cumulative richness of mammalian genera across cumulative annual samples from the Meier site.	148
4.3.	Cumulative richness of mammalian genera across cumulative annual samples from Cathlapotle.	149
4.4.	Relationship of mammalian genera richness and sample size per annual sample at the Meier site.	150
4.5.	Relative abundances of fifteen size classes of mollusk shells recovered during hand picking from the excavation, and recovered from fine-mesh sieves.	153
4.6.	The effect of passing sediment through screens or sieves on recovery of mammal remains relative to hand picking specimens from an excavation unit.	153
4.7.	Cumulative percentage recovery of remains of different size classes of mammals.	156
4.8.	Model of the relationship between area sampled and number of taxa identified.	160
4.9.	Two models of the results of rarefaction.	161
4.10.	Rarefaction curve and 95 percent confidence intervals of richness of mammalian genera based on six annual samples from the Meier site.	166
4.11.	Examples of perfectly nested faunas and poorly nested faunas.	168
4.12.	Nestedness diagram of eighteen assemblages of mammalian genera from eastern Washington State.	170
5.1.	Two fictional faunas with identical taxonomic richness values but different taxonomic evenness.	176
5.2.	Three fictional faunas with varying richness values and varying evenness values.	177
5.3.	Relationship between genera richness and sample size in eighteen mammalian faunas from eastern Washington State.	182
5.4.	Relationships between NISP and NTAXA of small mammals per stratum at Homestead Cave, Utah.	182
5.5.	Relationship between NISP and NTAXA per stratum at Le Flageolet I, France.	184
5.6.	Two Venn diagrams based on the Meier site and Cathlapotle site collections.	187
5.7.	Bivariate scatterplot of relative abundances of mammalian genera at the Meier site and Cathlapotle.	190
5.8.	Rarefaction analysis of eighteen assemblages of mammal remains from eastern Washington State.	191

5.9.	The relationship between taxonomic heterogeneity and sample size in eighteen assemblages of mammal remains from eastern Washington State.	194
5.10.	Frequency distribution of NISP values across six mammalian genera in a collection of owl pellets.	195
5.11.	Relationship between taxonomic evenness and sample size in eighteen assemblages of mammal remains from eastern Washington State.	197
5.12.	Relationship between sample size and the reciprocal of Simpson's dominance index in eighteen assemblages of mammal remains from eastern Washington State.	198
5.13.	Percentage abundance of deer in eighteen assemblages from eastern Washington State.	202
5.14.	Abundance of bison remains relative to abundance of all ungulate remains over the past 10,500 years in eastern Washington State.	203
5.15.	Bivariate scatterplot of elk abundances relative to the sum of all ungulate remains in eighty-six assemblages from eastern Washington State.	206
5.16.	Bivariate scatterplot of elk–deer index against stratum at Emeryville Shellmound.	208
5.17.	Relative abundances of <i>Neotoma cinerea</i> and <i>Dipodomys</i> spp. at Homestead Cave.	210
5.18.	Bivariate scatterplot of elk abundances relative to the sum of all ungulate remains in eighty-six assemblages from eastern Washington State summed by age for consecutive 500-year bins.	212
6.1.	Relationship of NISP and MNE values for deer remains from the Meier site.	225
6.2.	Relationship of NISP and MNE values for wapiti remains from the Meier site.	225
6.3.	Frequency distributions of NISP and MNE abundances per skeletal part for deer remains from the Meier site.	226
6.4.	Frequency distributions of NISP and MNE abundances per skeletal part for wapiti remains from the Meier site.	227
6.5.	Frequency distribution of skeletal parts in single skeletons of four taxa.	229
6.6.	Comparison of MNE of left skeletal parts and MNE of right skeletal parts in a collection of pronghorn bones.	231

6.7.	Frequencies of skeletal elements per category of skeletal element in a single artiodactyl carcass.	234
6.8.	MNE and MAU frequencies for a fictional data set.	235
6.9.	MNE values plotted against the MNE skeletal model.	236
6.10.	MAU values plotted against the MAU skeletal model.	237
6.11.	MAU values for two collections with different MNI values.	240
6.12.	%MAU values for two collections with different MNI values.	241
6.13.	Relationship between Shotwell's CSI per taxon and NISP per taxon for the Hemphill paleontological mammal assemblage.	243
6.14.	Relationship between Thomas's CSI per taxon and NISP per taxon for the Smoky Creek zooarchaeological mammal collection.	245
6.15.	Bar graph of frequencies of skeletal parts for two taxa.	247
6.16.	Model of the relationship between fragmentation intensity and NISP.	253
6.17.	Model of the relationship between NISP and MNE.	254
6.18.	Relationship between NISP and MNE values for size class II cervids and bovids at Kobeh Cave, Iran.	257
6.19.	Relationship between NISP and MNE values for saiga antelope at Prolom II Cave, Ukraine.	258
6.20.	Relationship between NISP and MNE values for caprine remains from Neolithic pastoral site of Ngamuriak, Kenya.	260
6.21.	Relationship between $\sum(\text{lefts} + \text{rights})$ and MNI per skeletal part.	262
7.1.	Weathering profiles for two collections of ungulate remains from Olduvai Gorge.	269
7.2.	Relationship between years since death and the maximum weathering stage displayed by bones of a carcass.	271
7.3.	Weathering profiles based on fictional data for a collection of bones with skyward surfaces representing one profile and groundward surfaces representing another profile.	272
7.4.	Frequency distribution of seven classes of burned bones in two kinds of archaeological contexts.	275
7.5.	Relationship between number of arm strokes and number of cut marks on thirty-one skeletal elements.	291
7.6.	Relationship between number of arm strokes necessary to deflesh a bone and the amount of flesh removed.	293
7.7.	Relationship between number of cut marks and the amount of flesh removed from thirty-one limb bones.	293

7.8.	Relationships between number of cut marks and the amount of flesh removed from six hindlimbs in each of three carcass sizes.	295
7.9.	Relationship between number of cut marks and the amount of flesh removed from eighteen hindlimbs.	295
8.1.	Relationship between NISP and MNI in seven paleontological assemblages of bird remains from North America.	304
8.2.	Relationship between NISP and MNI in eleven paleontological assemblages of mammal remains from North America.	304
8.3.	Relationship between NISP and MNI in twenty-two archaeological assemblages of bird remains from North America.	305
8.4.	Relationship between NISP and MNI in thirty-five archaeological assemblages of mammal remains from North America.	306

# LIST OF TABLES

---

1.1.	An example of the Linnaean taxonomy.	page 6
1.2.	Fictional data on the absolute abundances of two taxa in six chronologically sequent strata.	15
1.3.	Description of the mammalian faunal record at Meier and at Cathlapotle.	19
2.1.	Fictional data on abundances of three taxa in five strata.	32
2.2.	Data in Table 2.1 adjusted as if each individual of taxon A had ten skeletal elements per individual, taxon B had one skeletal element per individual, and taxon C had five skeletal elements per individual.	33
2.3.	The differential exaggeration of sample sizes by NISP.	36
2.4.	Some published definitions of MNI.	40
2.5.	A fictional sample of seventy-one skeletal elements representing a minimum of seven individuals.	45
2.6.	Statistical summary of the relationship between NISP and MNI for mammal assemblages from Meier and Cathlapotle.	53
2.7.	Statistical summary of the relationship between NISP and MNI for mammal assemblages from fourteen archaeological sites in eastern Washington State.	55
2.8.	Maximum distinction and minimum distinction MNI values for six genera of mammals in a sample of eighty-four owl pellets.	55
2.9.	Adams's data for calculating MNI values based on <i>Odocoileus</i> sp. remains.	57
2.10.	Differences in site total MNI between the MNI minimum distinction results and the MNI maximum distinction results.	59
2.11.	The most abundant skeletal part representing thirteen mammalian genera in two (sub)assemblages at Cathlapotle.	62

2.12. Fictional data showing how the distribution of most abundant skeletal elements of one taxon can influence MNI across different aggregates.	63
2.13. Fictional data showing how the distribution of skeletal elements of two taxa across different aggregates can influence MNI.	64
2.14. Fictional data showing that identical distributions of most common skeletal elements of two taxa across different aggregates will not influence MNI.	65
2.15. Ratios of abundances of pairs of taxa in eighty-four owl pellets.	72
3.1. Biomass of deer and wapiti at Cathlapotle.	86
3.2. Meat weight for deer and wapiti at Cathlapotle, postcontact assemblage.	90
3.3. Comparison of White's conversion values to derive usable meat with Stewart and Stahl's conversion values to derive usable meat.	91
3.4. Variation by age and sex of wapiti butchered weight as a percentage of live weight.	92
3.5. Weight of a 350-kilogram male wapiti in various stages of butchering.	93
3.6. Descriptive data on animal age, bone weight per individual, and soft-tissue weight per individual domestic pig.	97
3.7. Statistical summary of the relationship between bone weight and weight of various categories of soft-tissue for domestic pig.	98
3.8. Descriptive data on dry bone weight per anatomical portion and total weight per anatomical portion for domestic sheep.	100
3.9. Statistical summary of the relationship between bone weight and gross weight or biomass of skeletal portions of two domestic sheep.	101
3.10. Relationship between NISP and bone weight of mammalian taxa in seventeen assemblages.	104
3.11. Results of applying the bone-weight allometry equation to five randomly generated collections of domestic sheep bone.	105
3.12. Deer astragalus length and live weight.	110
3.13. Ubiquity and sample size of twenty-eight mammalian taxa in eighteen sites.	116
3.14. Ubiquity and sample size of mammalian taxa across analytical units in two sites.	118
3.15. Fictional data illustrating influences of NISP and the number of pairs on the Lincoln–Petersen index.	126
3.16. Abundances of beaver and deer remains at Cathlapotle, and WAE values and ratios of NISP and WAE values per taxon per assemblage.	135

3.17. Estimates of individual body size of seventeen white-tailed deer based on the maximum length of right and left astragali.	139
4.1. Volume excavated and NISP of mammals per annual field season at the Meier site.	144
4.2. Annual NISP samples of mammalian genera at the Meier site.	145
4.3. Annual NISP samples of mammalian genera at Cathlapotle.	147
4.4. Mammalian NISP per screen-mesh size class and body-size class for three sites.	155
4.5. Two sets of faunal samples showing a perfectly nested set of faunas and a poorly nested set of faunas.	169
5.1. Sample size, taxonomic richness, taxonomic heterogeneity, taxonomic evenness, and taxonomic dominance of mammalian genera in eighteen assemblages from eastern Washington State.	181
5.2. $\sum$ NISP and NTAXA for small mammals at Homestead Cave, Utah.	183
5.3. $\sum$ NISP and NTAXA for ungulates at Le Flageolet I, France.	184
5.4. NISP per taxon in two chronologically distinct samples of eighty-four owl pellets.	188
5.5. Expected values and interpretation of taxonomic abundances in two temporally distinct assemblages of owl pellets.	188
5.6. Derivation of the Shannon—Wiener index of heterogeneity for the Meier site.	193
5.7. Total NISP of mammals, NISP of deer, and relative abundance of deer in eighteen assemblages from eastern Washington State.	199
5.8. Frequencies of bison and nonbison ungulates per time period in ninety-one assemblages from eastern Washington State.	204
5.9. Frequencies of elk, deer, and medium artiodactyl remains per stratum at Emeryville Shellmound.	207
5.10. Frequencies of two taxa of small mammal per stratum at Homestead Cave.	209
6.1. MNE values for six major limb bones of ungulates from the FLK <i>Zinjanthropus</i> site.	219
6.2. Fictional data showing how the distribution of specimens of two skeletal elements across different aggregates can influence MNE.	223
6.3. NISP and MNE per skeletal part of deer and wapiti at the Meier site.	224
6.4. Frequencies of major skeletal elements in a single mature skeleton of several common mammalian taxa.	228
6.5. MNE frequencies of left and right skeletal parts of pronghorn from site 39FA83.	230

6.6.	Expected MNE frequencies of pronghorn skeletal parts at site 39FA83, and adjusted residuals and probability values for each.	232
6.7.	Frequencies of skeletal elements in a single generic artiodactyl skeleton.	233
6.8.	MNE and MAU frequencies of skeletal parts and portions.	236
6.9.	MAU and%MAU frequencies of bison from two sites.	239
6.10.	Skeletal-part frequencies for two taxa of artiodactyl.	246
6.11.	Expected frequencies of deer and wapiti remains at Meier, adjusted residuals, and probability values for each.	249
6.12.	Ratios of NISP:MNE for four long bones of deer in two sites on the coast of Oregon State.	252
6.13.	African bovid size classes.	255
6.14.	NISP and MNE frequencies of skeletal parts of bovid/cervid size class II remains from Kobeh Cave, Iran.	256
6.15.	NISP and MNE frequencies of skeletal parts of saiga antelope from Prolom II Cave, Ukraine.	257
6.16.	Relationship between NISP and MNE in twenty-nine assemblages.	259
6.17.	NISP and MNI frequencies of skeletal parts of caprines from Ngamuriak, Kenya.	260
6.18.	Relationship between NISP and MNI per skeletal part or portion in twenty-two assemblages.	261
7.1.	Weathering stages as defined by Behrensmeyer.	267
7.2.	Weathering stage data for two collections of mammal remains from Olduvai Gorge.	268
7.3.	Expected frequencies of specimens per weathering stage in two collections, adjusted residuals, and probability values for each.	269
7.4.	Frequencies of cut marks per anatomical area on six experimentally butchered goat hindlimbs.	288
7.5.	Frequencies of arm strokes and cut marks on sixteen limbs of cows and horses.	290
7.6.	Number of cut marks generated and amount of meat removed from eighteen mammal hindlimbs by butchering.	294
8.1.	Statistical summary of relationship between NISP and MNI in collections of paleontological birds, paleontological mammals, archaeological birds, and archaeological mammals.	303

# PREFACE

---

Several years ago I had the opportunity to have a relaxed discussion with my doctoral advisor, Dr. Donald K. Grayson. In the course of that discussion, I asked him if he would ever revise his then 20-year-old book titled *Quantitative Zooarchaeology*, which had been out of print for at least a decade. He said “No” and explained that the topic had been resolved to his satisfaction such that he could do the kinds of analyses he wanted to do. A spur-of-the-moment thought prompted me to ask, “What if I write a revision?” by which I meant not literally a revised edition but instead a new book that covered some of the same ground but from a 20-years-later perspective. Don said that he thought that was a fine idea.

After the conversation with Grayson, I began to mentally outline what I would do in the book. I realized that it would be a good thing for me to write such a book because, although I thought I understood many of the arguments Grayson had made regarding the counting of animal remains when I was a graduate student, there were other arguments made by other investigators subsequent to the publication of Grayson’s book that I didn’t know (or if I knew of those arguments, I wasn’t sure I understood them very well). I also knew that the only way for me to learn a topic well was to write about it because such a task forced me to learn its nuances, its underpinning assumptions, the interrelations of various aspects of the argument, and all those things that make an approach or analytical technique work the way that it does (or not work as it is thought to, as the case may be).

As I mentally outlined the book over the next several months, it occurred to me that at least one new quantitative unit similar to the traditional ones Grayson had considered had become a focus of analytical attention over the two decades subsequent to the publication of Grayson’s book (MNE, and the related MAU). And an increasing number of paleozoologists were measuring taxonomic diversity – a term that had several different meanings for several different variables as well as being measured several different ways. What were those measurement techniques and

what were those measured variables? Finally, there were other kinds of phenomena that zooarchaeologists and paleontologists had begun to regularly tally and analyze. These phenomena – butchering marks, carnivore gnawing marks, rodent gnawing marks, burned bones – had become analytically important as paleozoologists had come to realize that to interpret the traditional quantitative measures of taxonomic abundances, potential biases in those measures caused by differential butchery, carnivore attrition, and the like across taxa had to be accounted for. As I indicate in this volume, there are several ways to tally up carnivore gnawing marks and the like, and few analysts have explored the fact that each provides a unique result.

Finally, it had become clear to me during the 1990s that many paleozoologists were unaware of what I took to be two critical things. First, zooarchaeologists seemed to seldom notice what is published in paleontological journals; at least they seldom referenced that literature. Thus, they were often ignorant of various suggestions made by paleontologists regarding quantitative methods. Paleontologists were equally unaware of what zooarchaeologists have determined regarding quantification of bones and shells and teeth. Therefore, it seemed best to title this volume *Quantitative Paleozoology* for the simple reason that were it to be titled “Quantitative Zooarchaeology,” it likely would not be read by paleontologists. A very interesting book with the title *Quantitative Zoology* coauthored by a paleontologist (Simpson et al. 1960) already exists, so that title could not be used, aside from it being misleading. *Quantitative Paleozoology* is a good title for two reasons. The first reason is that the subject materials, whether collected by a paleontologist or an archaeologist, do not have a proximate zoological source (though their source is ultimately zoological) but rather have a proximate geological source, whether paleontological (without associated human artifacts) or archaeological (with associated and often causally related human artifacts). I conceive of all such remains as paleozoological. The second reason *Quantitative Paleozoology* is a good title is that the volume concerns how to count or tally, how to quantify zoological materials and their attributes, specifically those zoological remains recovered from geological contexts. Not all such topics are discussed here, but many are; for an introduction to many of those that are not, see Simpson et al. (1960), a still-useful book that was, fortunately, reprinted in 2003.

The second critical thing that many paleozoologists seem to be unaware of is basic statistical concepts and methods. I was stunned in 2004 to learn that an anonymous individual who had reviewed a manuscript I submitted for publication did not know what a “closed array” was and therefore did not understand why my use of this particular analytical tool could have been influencing (some might say biasing, but that is a particular kind of influencing) the statistical results. In the 1960s and early 1970s, many archaeologists and paleontologists did not have very high levels of statistical sophistication; I had thought that most of them did have such sophistication (or at

least knowledge of the basics) in the twenty-first century. The anonymous reviewer's comments indicate that at least some of them do not. Therefore, it seemed that any book on quantitative paleozoology had to include brief discussions of various statistical and mathematical concepts. In order to not dilute the central focus of the volume – quantitative analysis of paleozoological remains – I have kept discussion of statistical methods to a minimum, assuming that the serious reader will either already know what is necessary or will learn it as he or she reads the book. I have, however, devoted the first chapter to several critical mathematical concepts as well as some key paleozoological concepts.

Many of the faunal collections used to illustrate various points in the text were provided over the years by friends and colleagues who entrusted me with the analysis of those collections. Many of the things I have learned about quantitative paleozoology are a direct result of their trust. To these individuals, I offer my sincere thanks: Kenneth M. Ames, David R. Brauner, Jerry R. Galm, Stan Gough, Donald K. Grayson, David Kirkpatrick, Lynn Larson, Frank C. Leonhardy, Dennis Lewarch, Michael J. O'Brien, Richard Pettigrew, and Richard Ross. Perhaps more importantly, any clarity this book brings to the issues covered herein is a result of the collective demand for clarity by numerous students who sat through countless lectures about the counting units and methods discussed in this book. A major source of inspiration for the first several chapters was provided in 2004 by the Alaska Consortium of Zooarchaeologists (ACZ). That group invited me to give a daylong workshop on the topics of quantification and taphonomy, and that forced me to think through several things that had previously seemed less than important. I especially thank Diane Hansen and Becky Saleeby of the ACZ for making that workshop experience memorable.

An early draft of the manuscript was reviewed by Jack Broughton, Corey Hudson, Alex Miller, and an anonymous individual. Broughton and the anonymous reviewer ensured that a minimum of both glaring errors in logic and stupid errors in mathematics remain in this version. Broughton and the anonymous reviewer insisted that I include several recently described analytical techniques, and they identified where I overstepped and where I misstepped. These individuals deserve credit for many of the good things here.

I wrote much of the first draft of this volume between July 2005 and August 2006. During that time, I lost my younger brother and both parents. They all had an indirect hand in this book. My parents taught me to hunt and fish, and all of the things that accompany those activities. My brother did not discourage me from collecting owl pellets from his farm equipment shed, or laugh too hard when I collected them; he even grew to appreciate what could be learned from the mouse bones they contained. I miss them all, and I dedicate this book to the three of them.

June 2007



# Tallying and Counting: Fundamentals

Early in the twentieth century, paleontologist Chester Stock (1929) was, as he put it, faced with “recording a census” of large mammals from the late Pleistocene as evidenced by their remains recovered “from the asphalt deposits of Rancho La Brea,” in Los Angeles, California. Paleontologist Hildegarde Howard (1930) was faced with a similar challenge with respect to the bird remains from Rancho La Brea. Stock and Howard could have merely listed the species of mammals and the species of birds, respectively, that were represented by the faunal remains they had – they could have constructed an *inventory* of taxa – but they chose to do something more informative and more analytically powerful. They tallied up how many individuals of each species were represented by the remains – they each produced a census. The quantitative unit they chose became known as the *minimum number of individuals*, or MNI, a unit that was quickly (within 25 years) adopted by many paleozoologists. We will consider this unit in some detail in Chapter 2, but here it is more important to outline how Stock and Howard defined it and why they decided to provide a census rather than an inventory of mammals and an inventory of birds.

Stock (1929:282) stated that the tally or “count” of each taxon was “determined by the number of similar parts of the internal skeleton as for example the skull, right ramus of mandible, left tibia, right scaphoid. In many cases the total number of individuals for any single group [read *taxon*] is probably a minimum estimate.” Howard (1930:81–82) indicated that “for each species, the left or the right of the [skeletal] element occurring in greatest abundance was used to make the count. . . . It is probable that in many instances the totals present a minimum estimate of the number of individuals [per taxon] actually represented in the collection.” We will explore why the procedure Stock and Howard used provides a “minimum” estimate of abundance in Chapter 2. Stock and Howard each produced a type of pie diagram to illustrate their respective censuses of mammalian and of avian creatures based on the bony remains of each (Figure 1.1).

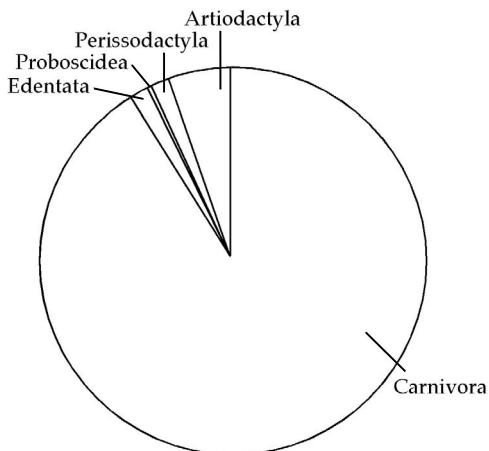


FIGURE 1.1. Chester Stock's pie diagram of abundances of five mammalian orders represented in faunal remains from Rancho La Brea. Redrawn from Stock (1929).

An inventory of the mammalian taxonomic orders Stock identified among the bones and teeth he studied would look like this:

- Carnivora
- Edentata
- Proboscidea
- Perissodactyla
- Artiodactyla

Clearly, the pie diagram in Figure 1.1 reveals more about the structure of the Rancho La Brea mammalian fauna because it contains not only the same set of taxonomic orders as the inventory, but it also contains measures of the abundances of animals belonging to each order. This example illustrates one of the major reasons why paleozoologists count or tally the animal remains they study. Taxa present in a collection can, on the one hand, be treated as attributes or as present or absent from a fauna, such as is given in the inventory above (sometimes referred to as a “species list” if that taxonomic level is used). On the other hand, abundances of each taxon provide a great deal more information about the prehistoric fauna. There are times when knowing only which taxa are present, or knowing only what the frequencies of different taxa are is all that is wanted or needed analytically. (Two faunas may have the same, or quite different, frequency distributions of individual organisms across taxa, and the research question may only require knowing the frequency distributions and not the taxa.) Knowing both, however, means we know more than when we know just one or the other. And that is a good reason to count faunal remains and to determine

a census. Counting faunal remains, particularly old or prehistoric remains, and the variety of attributes they display, whether the remains are from archaeological or paleontological contexts, is what this book is about.

There is already a book about counting animal remains recovered from archaeological and paleontological sites (Grayson 1984), and several other volumes cover some of the same ground, if in less detail (e.g., Hesse and Wapnish 1985; Klein and Cruz-Uribe 1984; Reitz and Wing 1999). Noting this, one could legitimately ask why another book on this topic is necessary. There are several reasons to write a new book. Much has happened in the field since Grayson (1984) published his book (and his book has been out-of-print for several years). Some of what has happened has been conceptually innovative, such as the definition of new quantitative units meant to measure newly conceived properties of the paleozoological record. Some of what has happened has been technically innovative, such as designing new protocols for tallying animal remains that are thought to provide more accurate reflections of what is represented by a collection of remains than tallies based on less technologically sophisticated methods. And, some of what has happened is misguided or archaic, such as arguing that if certain biological variables are not mathematically controlled for, then any count of taxonomic abundances is invalid. It is time (for these reasons) for a new, up-to-date examination of the quantitative units and counting protocols paleozoologists use in their studies.

There is yet another reason to produce a new book on quantitative paleozoology. Today, early in the third millennium, there are more people studying paleozoological collections than there were 20 years ago. These folks need to be able to communicate clearly and concisely with one another regarding their data and their analyses because the use of ambiguous terminology thwarts efficient communication and results in confusion. This point was made more than a decade ago with respect to the plethora of terms, many unfamiliar to those in the field, used for quantitative units in zooarchaeology (Lyman 1994a). Yet, the problem continues today. This problem had originally been identified more than 15 years earlier still by Casteel and Grayson (1977). For whatever reason, terminological ambiguity seems to plague paleozoology and continues to do so despite it being explicitly identified twice in the past 30 years.

In my earlier discussion of terminological ambiguity (Lyman 1994a), I did not advocate a particular terminology, nor am I doing so here. Clearly there are terms I prefer – the ones I use in this volume are the ones I learned as a student. What I am arguing here is that whatever terms or acronyms one uses, these must be clearly defined at the start so as to avoid misunderstanding. In reading and rereading the literature on quantitative paleozoology as I prepared this book, I was often dumbfounded when people used terms such as “bone” and “relative abundance” when it

was quite clear that they were discussing teeth and absolute abundances, respectively. Much of the remainder of this chapter is, therefore, devoted to terminology and definitions. For quick reference, I have included a glossary of key terms at the end of this volume.

In this introductory chapter, several basic mathematical and statistical concepts are defined. This is necessary because these concepts will be used throughout subsequent chapters and thus the concepts must be understood in order to follow the discussion in later chapters. Several basic paleozoological concepts are introduced and defined for the same reason. I begin with these concepts before turning to the mathematical and statistical concepts.

#### PALEOZOOLOGICAL CONCEPTS

Throughout this volume the focus is on vertebrates, especially mammals, because that is the taxonomic group which much of the literature concerns and because it is the group with which I am most familiar. However, virtually every thing that is said about quantifying vertebrate remains and their attributes holds with equal force for invertebrates (e.g., Claassen 1998:106–107).

In many discussions of how paleofaunal remains are tallied, and even in some discussions of how modern animal bones should be counted, the reader may encounter the term “skeletal element.” Or, one might encounter the term “bone,” or “tooth,” or “shell,” or any of many other similar, more or less synonymous general terms for skeletal remains. But if one collection comprises ten “bones” of a skeleton and another consists of eleven “bones” of another skeleton of the same species as the first, is the latter more anatomically complete than the former? Is the taxon less abundant in the first collection than in the second? If you think the answer is “Yes” to either question, you might be correct. But you could be wrong if when the analyst tallied specimens no distinction was made between anatomically complete bones and fragments of bones. The lesson is simple. If we are going to tally up skeletal parts and want to compare our tally with that of another analyst working with another collection, we had best be sure that we counted skeletal parts the same way that the other person did. What, then, exactly is a skeletal element?

Paleontologist Michael Voorhies (1969:18) distinguished between “fragments” and “elements or bones,” but we need something more explicit and inclusive because not all skeletal elements are, technically, bones. Some are teeth, some are horns, and some are antlers, and so on. Following Arnold Shotwell (1955, 1958), Donald Grayson (1984) and Catherine Badgley (1986) provide useful terminology and definitions. A

*skeletal element* is a complete discrete anatomical unit such as a bone, tooth, or shell. The critical phrase is *complete discrete anatomical unit*. Each such item is a discrete “anatomical organ” (Francillon-Vieillot et al. 1990:480) that does not lose its integrity or completeness when it is removed from an organism. A humerus, a tibia, a carpal, a first lower molar – each is a skeletal element. One might correctly note that “discreteness” depends on the age or ontogenetic stage of development of the organism, but many paleozoologists would not tally the proximal epiphysis of a humerus and the diaphysis of that humerus as two separate specimens if it was clear that the two specimens went together (an issue we return to in Chapter 2). Those same paleozoologists usually don’t tally up each individual tooth firmly set in a mandible, along with the dentary or mandible bone. These are potentially significant concerns but may ultimately be of minimal analytical import once we get into tallying specimens.

Not all faunal remains recovered from paleozoological deposits are anatomically complete; some are represented by only a part of the original skeletal element because of fragmentation. Thus, another term is necessary. A *specimen* is a bone, tooth, or shell, or fragment thereof. All skeletal elements are specimens, but not all specimens are skeletal elements. A distal humerus, a proximal tibia, and a fragment of a premolar are all specimens that derive from skeletal elements; phenomenologically they are not, technically, anatomically complete skeletal elements. *Specimen* is an excellent term for many counting operations because it is value-free in the sense that it does not reveal whether specimen A is anatomically more complete, or less complete, than specimen B. We can record whether specimen A is anatomically complete, and if it isn’t, we can record the portion of a complete element that is represented by a fragment, if our research questions demand such. *Specimen* is also a better generic term than *skeletal element* for the individual skeletal remains we study because *skeletal element* implies that a complete anatomical unit is represented. The problem with the terms “bone” and “tooth” and the like are that sometimes when analysts use them they mean both anatomically complete skeletal elements as defined here and incomplete skeletal elements. Failure to distinguish the two kinds of units – skeletal element and specimen – can render separate tallies incomparable and make the significance of various analyses obscure. Throughout this volume, I use the term *skeletal part* as a synonym for *specimen*, but whereas the latter is a general category that can include many and varied anatomical portions, skeletal part is restricted to a particular category of anatomical portion, say, distal humerus. *Skeletal portion* is sometimes used in the same category-specific way that skeletal part is but will usually mean a multiple skeletal element segment of a skeleton, such as a forelimb.

Henceforth, in this volume, *specimen* will be used to signify any individual skeletal remain, whether anatomically complete or not. Unfortunately, the terms “skeletal

Table 1.1. An example of the Linnaean taxonomy

Taxonomic level	Taxonomic name	Common name
Kingdom	Animalia	Animals
Phylum	Vertebrata	Vertebrate
Class	Mammalia	Mammals
Infraclass	Eutheria	Placental mammal
Order	Carnivora	Carnivores
Family	Canidae	Canids
Genus	<i>Canis</i>	Dogs, coyotes, wolves, and allies
Species*	<i>latrans</i>	coyote

\*Technically, the species name is *Canis latrans*; *latrans* is the specific epithet.

“element” and “element” are still often used to denote anatomically incomplete items. An effort is made throughout this book to make clear what exactly is being tallied and how it is being tallied. In this respect, what are usually tallied are what are termed “identified” or “identifiable” specimens. Typically, this means identified as to biological taxon, usually genus or species, represented by a bone, tooth, or shell (Driver 1992; Lyman 2005a). To identify skeletal remains, one must know the structure of the Linnaean taxonomy, an example of which is given in Table 1.1. One must also know the basics of skeletal anatomy, by which is meant that one must know the difference between a scapula and a radius, a femur and a cervical vertebra, a clavicle and a rib, and so on. Finally, the person doing the identifications must be able to distinguish intertaxonomic variation from intrataxonomic variation. *Intrataxonomic variation* is also sometimes termed “individual variation” within the species level of the taxonomy. I presume that readers of the book know these things, along with anatomical location and direction terms used in later chapters.

The importance of the requirements for identification should be apparent when one realizes that “identification” involves questions such as: Is one dealing with a mammal or a bird? If it is a mammal, is it a rodent or a carnivore? If it is a carnivore, is it a canid, a felid, a mustelid, or any of several other taxa of carnivores? The importance of the other knowledge requirement – basic skeletal anatomy – will assist in answering the questions just posed. The importance of distinguishing intertaxonomic from intrataxonomic variation is usually (and best) met by consultation of a comparative collection of skeletons of known taxonomic identity. The procedure is simple. Compare the taxonomically unknown paleozoological specimen with comparative specimens of known taxonomy until the best match is found. Often the closest match will be obvious, and the unknown specimen is “identified” as belonging to the same taxon as the known comparative specimen. Sometimes this means that

one may be able to determine the species represented by the paleozoological specimen, but other times only the genus or perhaps only the taxonomic family or order will be distinguishable.

Taxonomic identification is a complex matter that is discussed at length in other contexts (e.g., Driver 1992; Lyman 2005a; and references therein). Blind tests of identification results (e.g., Gobalet 2001) highlight the practical and technical difficulties. For one thing, what is “identifiable” to one analyst may not be to another (e.g., Grayson 1979). Gobalet (2001) provides empirical evidence for such interanalyst variation. It is precisely because of such interobserver differences and the interpretive significance of whether, say, a bone is from a bobcat (*Lynx rufus*) or a North American lynx (*Lynx canadensis*) that paleontologists developed a standardized format for reporting their results. Specimens (not necessarily anatomically incomplete skeletal elements) are illustrated and are verbally described with taxonomically distinctive criteria highlighted so that other paleontologists can independently evaluate the anatomical criteria used to make the taxonomic identification. Zooarchaeologists have been slow to understand the importance of this reporting form (see Driver [1992] for a noteworthy exception). This is not the place to delve further into the nuances of taxonomic identification and how to report and describe identified specimens. What is important here is to note that skeletal remains – faunal specimens – are usually tallied by taxon. “There are X remains of bobcats and Y remains of lynx.” So, identification must precede tallying. To make taxonomic identifications, one must first determine which skeletal element is represented by a specimen in order to know whether the paleozoological unknown should be compared to femora, humeri, tibiae, and so on. And sometimes the frequencies of each skeletal element or each part thereof are analytically important.

The final paleozoological concept that requires definition is *taphonomy*. The term was originally coined by Russian paleontologist I. A. Efremov (1940:85) who defined it as “the study of the transition (in all details) of animal remains from the biosphere into the lithosphere.” Although not without precedent, Efremov’s term is the one paleontologists (and an increasing number of paleobotanists) use to refer to the processes that influence the creation and preservation (or lack thereof) of the paleobiological record. We will have reason to return again and again to this basic concept; here it suffices to note that a taphonomic history concerns the formation of an assemblage of faunal remains. Such a history begins with the accumulation and deposition of the first specimen, continues through the deposition of the last specimen, through the preservation, alteration, and destruction of remains, and up to collection of a sample of the remains by the paleontologist (see Lyman [1994c] for more complete discussion). Along the way, faunal remains are modified, broken, and even destroyed. The modification, fracture, and destruction processes create

and destroy different kinds of phenomena the observation of which can generate quantitative data.

A final note about how paleozoological data are presented in the book. Capital letters are used to denote upper teeth, lower case letters to denote lower teeth, and a lowercase d to denote deciduous premolars. Thus, a permanent upper second premolar is P<sub>2</sub>, a deciduous lower third premolar is dp<sub>3</sub>, and a lower first molar is m<sub>1</sub>. The capital letter L is used to signify the left element of bilaterally paired bones, and the capital letter R is used to signify the right element. In general, D stands for distal, and P stands for proximal. The critical thing to remember is the difference between a *specimen* and a *skeletal element*; both terms will reappear often in what follows, and both kinds of units can be identified and tallied.

#### MATHEMATICAL AND STATISTICAL CONCEPTS

This book is about quantification, but the topics covered include different sorts of quantification, particularly counting or tallying units, methods of counting, and analyzing counts. A term that might have been used in the title of the book, were it not for its generality, is *measurement*. Typically this term is defined as assigning a numerical value to an observation based on a rule governing the assignment. The rule might be that length is measured in linear units of uniform size, such that we can say something like “Pencil A is 5 cm long and pencil B, at 10 cm of length, is twice as long as pencil A.” *Measurement* more generally defined concerns writing descriptions of phenomena according to rules. An *estimate* is a measurement assigned to a phenomenon (making a measurement) based on incomplete data. The process of *estimation* can involve judging how tall someone is in centimeters without the benefit of a tape measure, or studying a flock of birds and suggesting how many individuals there are without systematically tallying each one. Making estimates, like taking measurements, is a way to describe phenomena. Descriptions involve attributes of phenomena that may or may not have numerical symbols or values associated with them. Whether they do or not concerns what is often referred to as the scale of measurement of the attribute that is under scrutiny.

##### *Scales of Measurement*

Stock’s census of Rancho La Brea mammals (Figure 1.1) illustrates that quantitative data describing taxonomic abundances are more revealing than taxonomic

presence-absence data. Quantitative data often are subjected to a variety of mathematical manipulations and statistical analyses. Those manipulations and analyses are only valid if the data are of a certain kind. Four distinct scales of measurement are often distinguished (Blalock 1960; Shennan 1988; Stevens 1946; Zar 1996), and it is important that these be explicitly defined at the start because they will be referred to throughout the book.

*Nominal scales* of measurement are those that measure differences in kind. Of the several scales they contain the least amount of information. Numbers may be assigned to label nominal scale phenomena, such as 0 = male, 1 = female; or 11 = quarterback, 32 = fullback, and 88 = wide receiver on a football team. Or, numbers need not be assigned, but rather labels used such as Italian citizen, French citizen, and German citizen; or coyote (*Canis latrans*), wolf (*Canis lupus*), and domestic dog (*Canis familiaris*). Nominal scales of measurement do not include an indication of magnitude, ordering, or distance between categories, and are sometimes labeled *qualitative attributes* or *discontinuous variables*. They are qualitative because they record a phenomenon in terms of a quality, not a magnitude or an amount. They are discontinuous (or discrete) because it is possible to find two values between which no other intermediate value exists; there is (normally) no organism that is halfway between a male and a female within a bisexual species. Other scales of measurement tend to be quantitative because they specify variation more continuously. *Continuous variables* are those that can take any value in a series, and there is always yet another value intermediate between any two values. A tally of skeletal specimens of coyote in an archaeological collection may be 127 or 128, but there won't be a collection in which there are 127.5 or 127.3 or 127.924 specimens of coyote. But the lengths of coyote humeri are continuous; think about the numbers just noted as millimeters of length.

*Ordinal scales* of measurement are those that record greater than, less than relationships, but not the magnitude of difference in phenomena. They allow phenomena to be arranged in an order, say, from lesser to greater. "I am older than my children" is a statement of ordinal scale difference, as is "The stratum on the bottom of the stratigraphic column was deposited before the stratum on the top of the column" and "A year is longer than a month." There is no indication of the magnitude of difference in my age and the ages of my children, or in the length of time between the deposition of the bottom and top strata, nor in the duration of a year relative to the duration of a month. Instead, we only specify which phenomenon is older (or younger), or which was deposited first (or last), or which is longer in duration (or shorter). Sometimes when one uses an ordinal scale, measurements are said to be *relative* measurements because a measure of phenomenon A is made relative to

phenomenon B; A is older/shorter/heavier than B. Ordinal scale measurements may be (and often are said to be) *rank ordered* from greatest to least, or least to greatest, but the magnitude of distance between any two measurements in the ordering is unknown. Ordinal scale measurements are discrete insofar as there is no rank of “first and a half” between the rank of first and second (ignoring tied ranks).

*Interval scales* of measurement are those that record greater than or less than relationships and the magnitude of difference between phenomena. Both the order of measurements and the distance between them are known. My children are 23 and 25 years old; I am 56 years old, so I am 33 and 31 years older than my two children, respectively. The stratum on the bottom of the stratigraphic column has an associated radiocarbon date of 3000 BP and the stratum on top has an associated date of 500 BP, so the stratum on the bottom was deposited about 2,500 ( $^{14}\text{C}$ ) years before the stratum on top (assuming the dated materials in each stratum were formed and deposited at the same time as the strata were deposited). On average, a year is 365.25 days long whereas an average month is about 30.4 days in duration; the difference in duration of an average year and an average month is thus 334.85 days. The distance between 10 and 20 units (days, years, centimeters) is the same as the distance between 244 and 254 of those units, the same as the distance between 5337 and 5347 of those units, and so on. Interval scales are typically used to measure what are referred to as *quantitative variables*. Interval scale measurements, like ordinal scale ones, can be rank ordered from greatest to least, or least to greatest, but unlike with ordinal scale measures, the distance between any two interval scale measurements is known. Indeed it must be known else the variable is not interval scale. Interval scale measurements are generally continuous but may be discrete. If age is recorded only in whole years, then age is continuous but it is also discrete (ignoring for the sake of discussion that one might be 53.7 years old). Importantly, interval scale measures have an arbitrary zero point. It can be  $0^\circ\text{Celsius}$  outside, but there is still heat (if seemingly only a little) caused by the movement of molecules. The zero point on the Celsius scale is placed at a different location along the continuum of amount of molecular movement than is the zero point of the Fahrenheit temperature scale. Both zero points are arbitrary with respect to the amount of heat (molecular movement), thus both measures of temperature are interval scale.

*Ratio scales* of measurement are identical to interval scales but have a natural zero. Thus, the theoretical natural zero of temperature is  $-273^\circ\text{ Celsius}$  (or  $0\text{ Kelvin}$ , or  $-459^\circ\text{ Fahrenheit}$ ). There is no molecular movement at that temperature. Similarly, a mammal in a cage comprises 1 individual consisting of more than 100 bones and teeth, but if the cage is empty there are 0 (zero) individuals, 0 bones, and 0 teeth in the cage. Thus, if a taxon is represented by skeletal specimens in an assemblage, it is

absent; that is a natural zero. Essentially all quantitative measures in paleozoology – taxonomic abundances, frequency of gnawing damage, and so on – are potentially ratio scale. Whether they are in fact ratio scale or not is another matter.

Measurements of different scales allow (or demand, depending on your perspective) different statistical tests of different scales or power. Thus, ordinal scale measurements require ordinal scale statistical tests; interval/ratio scale measurements can be analyzed with either interval/ratio scale statistics or ordinal scale ones, but the reverse – applying interval/ratio scale statistics (or parametric statistics) to ordinal scale data – will likely violate various statistical assumptions.

Most paleozoologists working in the twentieth century sought ratio scale measures of the attributes of the ancient faunal remains they studied, just as Stock and Howard did. Although the optimism that such measures would eventually be designed has waned somewhat, there are still many who hope for such, whether working with human remains (e.g., Adams and Konigsberg 2004), paleontological materials (e.g., Vermeij and Herbert 2004), or zooarchaeological collections (e.g., Marean et al. 2001; Rogers 2000a). We now know a lot more about taphonomy than we did even 20 years ago when Grayson (1984), Klein and Cruz-Uribe (1984), and Hesse and Wapnish (1985) noted that many problems with quantitative zooarchaeology originated in taphonomic histories. And we also know that many taphonomic analyses and interpretations of taphonomic histories require quantitative data and analyses of various sorts. Where taphonomy can influence quantitative paleozoology is noted throughout this volume, and it is occasionally suggested what we might do about those influences. The point here is that ratio scale measurements of faunal remains and many of their attributes may be precluded because of taphonomic history.

### *Measured and Target Variables: Reliability and Validity*

Other important statistical concepts concern the difference between a *measured variable* and a *target variable*. A measured variable is what we actually measure, say, how many gray hairs I have on my head. A target variable is the variable that we are interested in, say, my age. The critical question is this: Are the measured variable and the target variable the same variable, or are they different? If the latter, the question becomes: Are the two variables sufficiently strongly correlated that measuring one reveals something about the other? It is likely that the number of gray hairs on my head will be correlated with my age, assuming I do not artificially color my hair (either not gray, or gray). But although the color of the shirt I am wearing today

can be measured rather precisely, it is unlikely to indicate or correlate with my age (although the style of my shirt might).

The concepts of measured variable and target variable can be stated another way. When we measure something, are we measuring what we think we are measuring? Does the attribute we are measuring reflect the concept (e.g., length, age, color) we wish to describe (Carmines and Zeller 1979)? These questions serve to define the concept of *validity*. Is a radiocarbon age on a piece of burned wood a *valid* measure of the age of deposition of a fossil bone with which the wood is stratigraphically associated? Assuming no contamination of the sample of wood, and that the wood was deposited more or less simultaneously with the bone, it will be a valid measure if it derives from a plant that was alive at about the same time as the animal represented by the bone. Validity is a different property of a measurement than *reliability*, which simply defined means replicability, or, if we measure something twice, do we get the same answer? If, on the one hand, we measure the length of a femur today and get 12.5 cm, tomorrow we measure it and get 12.4 cm, and the next day we measure it and get 12.5 cm, then we are producing rather consistent and thus reliable measures of that femur's length. On the other hand, femur length is unlikely to be a valid measure of the time period when the represented animal was alive, regardless of the reliability of our measurements of length.

Another set of measurements will help underscore the significance of the preceding paragraph, and help highlight the differences between a *target variable* and a *measured variable*. A *fundamental measurement* (sometimes referred to as *primary data* [Clason 1972; Reitz and Wing 1999]) is one that describes an easily observed property of a phenomenon. Length of a bone, stage of tooth eruption in a mandible, and taxon represented by a shell are all fundamental measurements. A *derived measurement* (sometimes referred to as *secondary data*) is more complex than a fundamental one because it is based on multiple fundamental measurements. Derived measurements are defined by a specified mathematical (or other) relation between two or more fundamental measurements. A ratio of length to width exemplifies a derived measure. Derived measurements require analytical decisions above and beyond a choice of scale; do we calculate the ratio of length to width, or width to length, or width to thickness? As a result, derived measurements are sometimes difficult to relate clearly to theoretical or interpretive concepts. Derived measurements may nevertheless reveal otherwise obscure patterns in data even though relating those patterns to a target variable may be difficult.

The MNI measure mentioned above is the most widely known derived measurement in paleozoology. It depends on (i) tallies of (ii) each kind of skeletal element of (iii) each taxon in a collection, and often (iv) (but not always) other information, such as size of bones of a taxon. Each of the lower case Roman numerals denotes

a distinct fundamental measurement; each plays a role in deriving an MNI, as can several other fundamental measurements (considered in more detail in Chapter 2). A *flat or proxy measurement* will likely be more complex than either a fundamental or derived measurement because a flat measurement is more conceptual or abstract and less easily observed. The distinction of fundamental, derived, and proxy measurements is relevant to a measurement's accuracy. "Accuracy" refers to "the nearness of a measurement to the actual value of the variable being measured" (Zar 1996:5). Throughout this volume, major concerns are the accuracy and validity of derived measures or secondary data, and fundamental measures or primary data with respect to a target variable of interest. Does a particular derived measure, such as MNI, accurately reflect the abundance of individual organisms in a collection of bones and teeth (or shells)? Of organisms in a deposit? Of organisms on the landscape?

Stock's census of Rancho La Brea mammals was, he hoped, an accurate proxy measure of the structure and composition of the mammalian fauna on the landscape at the time of the deposition of the remains. That long-dead fauna is not directly visible or measurable, so how well the remains from the tar pits actually reflect or measure that fauna in terms of which taxon was most abundant and which was least abundant and a host of other properties (how accurately MNI measures the landscape fauna) cannot be determined. The validity of a flat or proxy measurement, or a measured variable, for reflecting a target variable of some sort is the key issue underpinning much of the discussion in this volume. This is so for the simple reason that many target variables in paleozoology cannot be directly measured reliably or validly with broken bones, isolated teeth, and fragments of mollusk shell. What this book is in part about is how well the measured variables and proxy measurements commonly used by paleozoologists measure or estimate the target variable(s) of interest. Two key questions to keep in mind throughout this book are: What is the target variable? How is the measured variable related to the target variable of interest? As a prelude to how important these questions are, think about this. Was Stock wise to use MNI (the derived and measured variable) to estimate the abundances of mammals *on the landscape* (the target variable) given that he only had animals that became mired in the pits of sticky tar at Rancho La Brea? Would he have been better off using, say, the tally of skulls (a different measured variable) to estimate the abundances of mammals *trapped in the tar pits* (a different target variable)?

### *Absolute and Relative Frequencies and Closed Arrays*

An *absolute frequency* is a raw tally of some set of entities, usually all of a particular kind. To note that there are ten rabbit bones and five turkey bones in a collection is

to note the absolute frequencies of specimens of each species. If one were to note that in that collection of fifteen specimens, 66.7 percent of the specimens were of rabbits and 33.3 percent were of turkey, then one would be noting the *relative frequency* of each species. Relative frequencies are termed such because they are *relative* to one another. A relative frequency is a quantity or estimate that is stated in terms of another quantity or estimate. The analyst could have different absolute abundances, say thirty rabbit bones and fifteen turkey bones, but rabbit bones would comprise the relative abundance of 66.7 percent of the collection and turkey bones would comprise 33.3 percent of that collection, the same as when there are ten rabbit bones and five turkey bones. Percentages and proportions of a total are relative frequencies. The term “relative frequencies” is sometimes used in the paleozoological literature to signify estimates in which a quantity is not stated but rather only that A is greater (or smaller, or less) than B. In such cases relative frequencies are equivalent to ordinal scales of measurement. In this volume, the term “relative frequencies” is used in the more typical sense of percentage or proportional abundances.

Relative frequencies are typically given as percentages of some total set of things, and the summed relative frequency is always 100 percent (proportions are fractions). When relative frequencies of kinds of things in a set of things are given as percentages, all of those frequencies must sum to 100 percent rather than 90 percent or 110 percent. Such percentage relative frequencies comprise what is called a *closed array* (proportions also form a closed array as they must sum to 1.0).

Another way to think about the difference between absolute and relative frequencies involves comparison of measurements. Let’s say we have two collections of faunal remains. In collection 1, taxon A is represented by 5 specimens and taxon B is represented by 10 specimens. In collection 2, taxon A is represented by 50 specimens and taxon B is represented by 55 specimens. The absolute difference in abundances of the two taxa in each collection is 5 specimens, but in collection 1, taxon A is only 50 percent as abundant as taxon B whereas in collection 2 taxon A is 90.9 percent as abundant as taxon B. Or, one could say that in collection 1 the relative abundances of taxa A and B are 33.3 percent and 66.7 percent, respectively, whereas the relative abundances of those taxa in collection 2 are 47.6 percent and 52.4 percent, respectively. The difference between absolute and relative frequencies is not a matter of which is correct and which is not, but rather they are simply two different ways to measure (describe) the frequencies of things.

Importantly, the absolute frequency of things of kind A in a collection will not change value if the absolute frequency of kind B in that collection changes, but the relative frequency of both A and B will change if the absolute frequency of either A or B changes. This last property is a characteristic – one could say diagnostic – of

*Table 1.2. Fictional data on the absolute abundances of two taxa in six chronologically sequent strata*

	Taxon A	Taxon B
Stratum VI	50 (71.4)	20 (28.6)
Stratum V	50 (62.5)	30 (37.5)
Stratum IV	50 (55.6)	40 (44.4)
Stratum III	50 (50.0)	50 (50.0)
Stratum II	50 (45.4)	60 (54.6)
Stratum I	50 (41.7)	70 (58.3)

Relative (percentage) abundances in parentheses.

closed arrays; they must sum to 100 percent. Consider the set of fictional data in Table 1.2. If we examine these data, we see that Taxon A does not change in absolute abundance over the stratigraphic sequence, but Taxon B does change in absolute abundance. However, relative abundance data suggest that both taxa change in abundance. This example reveals a final and critically important aspect of abundance data.

In paleozoology, absolute abundance data or raw tallies are often given, but when it comes to interpreting abundance data, it is in terms of relative abundances. Using the fictional data in Table 1.2 as an example, one might read something like the following:

Throughout the stratigraphic sequence (from stratum I as earliest or oldest to stratum VI as youngest) Taxon A increased in abundance relative to Taxon B, which decreased in relative abundance. Given that Taxon A prefers habitats that support vegetation adapted to cool-moist climatic conditions, and Taxon B prefers habitats indicative of warm-dry climatic conditions, then it seems that over the time span represented by strata I–VI, the local climate became progressively cooler and moister.

Notice that in the interpretation no mention is made of the absolute abundances of taxa A and B. Rather, their abundances *relative to each other* are the focus. The abundances are not even taken as measures of how many of either taxon was present on the landscape at the time the strata and faunal remains were deposited. Rather, the interpretation involves postulating a cause for the shift in relative abundances of two taxa. One could also postulate that hunting practices or procurement technology shifted, resulting in the shift in which taxon was taken more frequently. Resolving these sorts of issues is beyond the scope of this volume, but suffice it to say that regardless of the interpretive model one calls upon, *relative* abundances, in the case of this example relative *taxonomic* abundances, are interpreted.

## DISCUSSION

Quantitative data often comprise tallies of different kinds of phenomena. They might also include a set of measurements of, say, the length of individual specimens. This book concerns only the former kind of quantitative data – tallies. It is about how a paleozoologist might count phenomena (faunal specimens, or attributes thereof) when one seeks a measure of the magnitude of a particular variable that demands counts of phenomena (bones, teeth, shells, and fragments thereof, or burned bones, gnawed bones, or broken bones). How one chooses to tally those phenomena, and how those tallies are summarized and analyzed statistically, depend in large part on the research question asked and the target variable that one hopes to measure in order to answer that question. The choices likely will also depend on the presumed relationship of the target variable and the chosen measured variable. Discussion of how one determines the nature of that relationship in particular cases is beyond the scope of this volume. When necessary to assist discussions in later chapters, a particular relationship is assumed or identified.

The terms *assemblage* or *collection* denote an aggregate of faunal remains whose setness has been defined archaeologically (e.g., remains from an excavation unit), geologically (e.g., remains from a trash pit or a stratum), or analytically (e.g., all remains of a taxon). Graphs are used whenever possible to exemplify and illustrate concepts and analytical results, and to display relationships between variables. Statistical analyses are kept relatively simple and are used to evaluate particular properties of collections. In a few cases, statistical complexities are described in a clearly delineated box of text and may be skipped when reading the main content of a chapter. Data are often presented in table form so that the reader may replicate the statistical analyses (and graphs) to ensure understanding. This volume is, however, not meant to be exhaustive with respect to all of the myriad ways that faunal remains might be counted, or with respect to how the varied features faunal remains might display can be counted. Rather, most of the commonly used quantitative units (measured variables) and their attendant analyses serve as the background against which the discussion is framed. Target variables are identified and defined as necessary when discussing particular quantitative units.

This is not a book about taphonomic, zooarchaeological, or paleontological analyses. There are several excellent titles on each of these topics that are presently available (e.g., Lyman 1994c; Reitz and Wing 1999; Simpson et al. 1960, respectively). *Quantitative Paleozoology* is meant as a supplement to those other volumes because it covers in detail a limited range of topics relevant to various analytical methods and techniques described in each of those other volumes.

## BACKGROUND OF SOME FAUNAL SAMPLES

Throughout this volume, extensive use is made of data derived from actual zooarchaeological and paleontological collections of vertebrate, usually mammalian, remains. In several cases, the mammalian remains from a set of modern owl pellets collected in the 1990s are used to illustrate an analytical procedure or a concept (see Lyman and Lyman [2003] and Lyman et al. [2001, 2003] for more details on this collection). In the chapters that follow, many points are illustrated by analyzing faunas from various places and dating to various time periods. This helps to emphasize that many properties of the paleozoological record are in at least one sense *universal*, by which is meant that those properties are typically found in an *average* paleo-faunal assemblage. By “average” is meant *typical* and having multiple taxa (usually more than a half-dozen) and multiple identified specimens for the total collection (usually more than, say, 50 specimens). What are rather atypical if not rare or unusual are those collections that have hundreds if not thousands of specimens, all representing the same species. The well-known bison (*Bison* spp.) kill sites of North America do not seem particularly rare because publications on them are numerous, but in terms of the faunal record they are rather atypical. An even more unusual paleofauna would be one consisting of only a couple identified specimens (say, < 10), each representing a unique taxon. When necessary, these sorts of relatively unusual collections are mentioned, but otherwise typical faunas are used to illustrate quantitative concepts and analyses.

Using the same faunal samples throughout, it will be easy to track different kinds of interdependence, and how one analytical result influences whether or not another analysis is reasonable or even feasible. And, by the same token, if two faunal samples from basically the same geographic area and dating to the same time period are available, then other sorts of analytical insights can be gained. Thus two collections of mammal remains are used to illustrate significant points in later chapters. Analyses of the artifacts and features at each site are ongoing, so some of the background information is terse and incomplete. The lack of information on particular aspects of the collections will not make a difference to the points made in this volume.

The collections are zooarchaeological – they originate in an archaeological context. These are faunal remains that had associated artifacts; such assemblages are sometimes referred to as *archaeofaunas*. Both collections consist of mammal remains recovered from two late-prehistoric sites within 10 km of one another. Both sites are found in what is locally known as the Portland Basin or the Wapato Valley of northwestern Oregon state and southwestern Washington state. All mammalian remains from both sites were recovered from one-quarter-inch mesh screens in the field.

The Meier site (35CO5) is located downstream (north) of modern Portland, Oregon, on the floodplain of the Columbia River, on the Oregon side. The Meier site comprises a single large cedar-plank house that was occupied more or less continuously between approximately AD 1400 and AD 1800, and associated midden deposits (Ames 1996; Ames et al. 1992). The site was tested in 1973 and 1984. It underwent extensive excavations every year between 1987 and 1991, inclusively. The 1973 collection was made by Pettigrew (1981) and studied by Saleeby (1983). The 1984 test was directed by Ellis (n.d.); recovered faunal remains have not been analyzed. Kenneth Ames of Portland State University directed the excavations that took place in the late 1980s and early 1990s. I identified all mammalian remains collected by Ames during a 1993 research-sponsored leave when I worked with him in Portland.

The other site, Cathlapotle (45CL1), is located northeast of Meier, on the Washington side of the Columbia River, on a series of levees next to the river. The site was visited by Meriwether Lewis and William Clark in March of 1806 as they lead the Corps of Discovery eastward. At the time of their visit the site comprised several large cedar-plank houses and associated midden deposits (Ames et al. 1999; Ames and Maschner 1999:110). Radiocarbon dates indicate the main occupation began about AD 1450. Ceramic trade goods indicate that abandonment of the site occurred about AD 1834. Auger sampling of the Cathlapotle sediments took place in 1992–1993, and excavations took place each year from 1993 through 1996. Both the auguring and the excavations were under the direction of Ames. I identified all mammalian remains recovered from this site at the University of Missouri-Columbia campus.

The assemblages of mammalian remains from Meier and Cathlapotle were recovered from similar depositional contexts. At both sites, the deposits variously comprise exterior (midden and “yard”) deposits and interior deposits (inside of a house). Exterior deposits had very high organic content, lenses of fresh-water mussel shells, and other indications that they formed as primary or secondary dumps (Ames et al. 1999). Yard deposits are generally broad, sheet-like deposits that contain intact hearths, activity areas, pits, evidence of small structures, and so forth. They usually lack the very high organic content of middens though they can have organic content. Interior deposits were assigned to walls, benches (deposits below the 2 m-wide sleeping benches or platforms that ran along the interior side of the house walls), storage pits, and hearth areas. Faunal remains have not been sorted into these distinct depositional contexts as yet. Were they to be so assigned, it is likely that assemblages would be quite small. In later chapters, the influences of small sample sizes receive considerable attention.

The houses at Meier and Cathlapotle had extensive subfloor storage features that, at Meier at least, formed a cellar almost 2 m deep that extended under the house

Table 1.3. Description of the mammalian faunal record at Meier and at Cathlapotle

Taxon	Cathlapotle					
	Meier		Precontact		Postcontact	
	NISP	MNI	NISP	MNI	NISP	MNI
<i>Didelphis</i> *	—	—	—	—	10	1
<i>Scapanus</i>	14	4	—	—	3	2
<i>Sorex</i>	—	—	3	1	1	1
<i>Sylvilagus</i>	16	2	—	—	1*	1
<i>Lepus</i>	—	—	3	2	40	4
<i>Aplodontia</i>	5	1	61	8	57	8
<i>Tamias</i>	1	1	—	—	—	—
<i>Tamiasciurus</i>	2	1	—	—	—	—
<i>Thomomys</i>	9	5	—	—	—	—
<i>Castor</i>	329	9	111	5	238	7
<i>Peromyscus</i>	35	21	2	1	3	2
<i>Rattus</i> *	1	1	—	—	—	—
<i>Neotoma</i>	1	1	—	—	—	—
<i>Microtus</i>	100	41	10	4	55	22
<i>Ondatra</i>	337	13	56	7	36	4
<i>Erethizon</i>	1	1	—	—	—	—
<i>Canis</i>	90	7	18	3	17	2
<i>Vulpes</i>	2	1	4	1	—	—
<i>Ursus</i>	82	5	45	3	53	4
<i>Procyon</i>	272	20	109	6	84	8
<i>Martes</i>	19	4	1	1	1	1
<i>Mustela</i>	130	17	12	3	9	2
<i>Mephitis</i>	4	2	3	1	—	—
<i>Lutra</i>	45	3	28	3	26	4
<i>Puma</i>	9	1	5	1	5	2
<i>Lynx</i>	22	3	8	1	15	2
<i>Phoca</i>	40	3	26	2	34	2
<i>Ovis</i>	—	—	1	1	1	1
<i>Cervus</i>	832	10	1,091	12	1,793	24
<i>Odocoileus</i>	3,504	56	775	14	1,347	27
<i>Equus</i> *	—	—	—	—	4	1
<b>TOTAL</b>	<b>5,939</b>	—	<b>2,372</b>	—	<b>3,834</b>	—

\*Taxa are historically introduced and not native to the area, hence they are intrusive to site sediments.

floor between the sleeping platforms and the row of hearths in the house's center. The Cathlapotle features are less extensive, but are about 2 m wide by 2 m deep. They are below the sleeping platforms rather than next to them as at Meier. The mammalian remains from both sites were derived primarily from these storage pits and exterior areas.

Because, with few exceptions, the mammalian genera identified are monotypic (include only one species), the basic faunal identification and quantitative data are presented by genus (Table 1.3). Stratigraphy at Meier is extremely complex as a result of multiple episodes of house remodeling and rebuilding; temporally distinct assemblages could not be distinguished. This assemblage is, therefore, treated as a whole and not subdivided into subassemblages. The stratigraphy at Cathlapotle, on the other hand, though also rather complex as a result of the various things that people did at the site, was sufficiently clear that two temporally distinct (sub)assemblages of faunal remains could be distinguished. Many, but not all of the mammal remains from Cathlapotle could be sorted into a pre (Euroamerican) contact sample deposited between AD 1400 and AD 1792, and a postcontact sample deposited between AD 1792 and AD 1835. These two temporally distinct assemblages are referred to in later chapters simply as the precontact and postcontact assemblages. For most purposes the faunal remains that could not be assigned to a temporal period are not included in the analyses presented in later chapters.

The samples of identified remains from the two sites are not tremendously large by some standards (Table 1.3). However, based on data compiled by others (especially Casteel 1977, n.d.), both collections are of reasonable size. That they are not tremendously large is a benefit in the sense that this will assist with the detection of possible influences of sample size in later chapters. The three assemblages are described in Table 1.3. For additional information on the mammal collections not covered in later chapters, see Lyman (2004a, 2006b, 2006c), Lyman et al. (2002), Lyman and Ames (2004), and Lyman and Zehr (2003).

## Estimating Taxonomic Abundances: NISP and MNI

As paleontologists, Chester Stock and Hildegard Howard were interested in the abundance of the mammals that had walked the landscape and birds that had flown in the air above the landscape at the time the faunal remains from Rancho La Brea were deposited (Chapter 1). Zooarchaeologists (known as archaeozoologists in Europe), on the other hand, are typically interested in which taxa provided the most economic resources and which taxa provided little in the way of economic resources. Thus, as zooarchaeologist Dexter Perkins (1973:369) noted, “the primary objective of faunal analysis of material from an archaeological site [or from a paleontological site] is to establish the relative frequency of each species.” This target variable sought by paleontologists and zooarchaeologists concerns *taxonomic abundances*. What are the frequencies of the taxa in a collection?

In any given collection of paleozoological remains, one might wish to know if carnivores are less abundant than herbivores, just as they normally are on the landscape. Given what he knew about ecological trophic structure – that herbivores should outnumber carnivores – imagine Stock’s surprise to learn that the typically observed food pyramid or ecological trophic structure was upside down. The mammalian remains from Rancho La Brea represented more carnivores than herbivores – for a reason that many paleontologists thought was a taphonomic reason – because scavenging carnivores got “bogged down” or mired in the sticky tar seeping from the ground and failed to escape. Carnivores were abundant in that tar because they had died and become entombed there as a result of trying to exploit the carcasses of herbivores (and perhaps carnivore brethren) that had themselves become mired and subsequently died there.

The eminently sensible hypothesis that carnivore remains are more abundant than herbivores for taphonomic reasons (see Spencer et al. [2003] for a recent evaluation of this hypothesis) concerns the relationship between a target variable and a measured variable. Stock’s target variable was the frequencies of mammalian taxa comprising

the animal community on the landscape, but his measured variable was the sample of bones and teeth from the excavations of the tar pits. Taphonomists have developed an unwieldy terminology (see the glossary in Lyman 1994c), but several of the terms are useful here for keeping *target variables* and *measured variables* distinct. A *biocoenose* is a living community of organisms. Exactly what a community comprises is the subject of some debate. One definition provided by biologists is this: A *community* is comprised of “species that live together in the same place. The member species can be defined either taxonomically or on the basis of more functional ecological criteria, such as life form or diet” (Brown and Lomolino 1998:96).

But “most so-called communities are arbitrary and convenient segments of a continuum of species with overlapping ecological requirements, not involving a high level of interdependence” (Lawson 1999:7). Thus one commentator notes that a biological community can be defined one of two ways: As a group of organisms occupying a location, or as a group of organisms with ecological linkages among them (Southwood 1987). Often the focus is on the former at the expense of the latter; a community (which may include all organisms or a particular subset of organisms) is often defined by specific spatial boundaries (Magurran 1988:57). Perhaps not surprisingly, the “nature of boundaries separating adjacent communities is hotly disputed by community ecologists and paleoecologists” (Hoffman 1979:364). For the sake of discussion, we assume that a biological community and its boundaries can be defined on the landscape today.

The taxonomic composition and taxonomic abundances of a biocoenose might well be the target variable sought by a paleozoologist. However, that biocoenose is not what a paleozoologist studies. The organisms that comprise a biocoenose must die before a paleozoologist can study their mortal remains. A *thanatocoenose* is an assemblage of dead organisms; it is sometimes referred to as the death assemblage. Turner (1983) suggested that it be referred to as the “killed population,” but “killed” implies an active agent of death, such as a predator, disallowing death from other causes such as old age. Furthermore, the dead organisms may be a “population” in a statistical sampling sense, but they might not be, depending on the question asked. Those dead organisms may be a 100 percent sample of some set of dead organisms (e.g., all of those from a biocoenose, which, after all, must all die), but they may be less than that, such as when the set of dead organisms represents only part – a sample – of a biocoenose. In this case, the thanatocoenose is not, statistically speaking, a “population.” It is perhaps best, for this reason alone, to conceive of a thanatocoenose as a set of dead organisms, usually somehow stratigraphically or analytically bounded (see the discussion of a *faunule* later in this chapter).

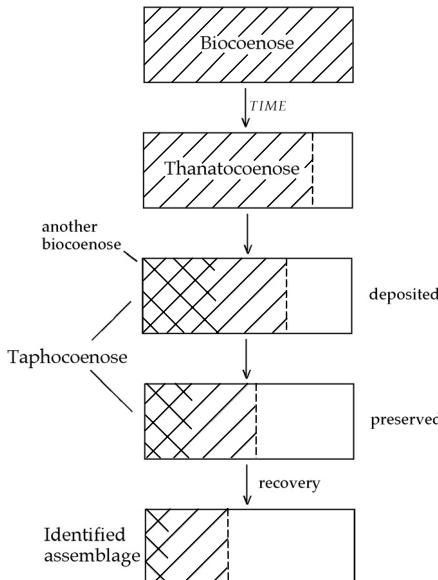


FIGURE 2.1. Schematic illustration of loss and addition to a set of faunal remains studied by a paleozoologist.

Given that paleozoologists sample the geological record (i.e., where faunal remains are deposited as a particular kind of sedimentary particle), they don't always have a complete thanatocoenose lying on the lab table. Furthermore, the organisms whose remains comprise a thanatocoenose may derive from one community (or biocoenose) or they may derive from more than one community (Shotwell 1955, 1958). A *taphocoenose* is the set of remains of organisms (in our case, faunal remains) found buried (or perhaps exposed) and spatially (usually stratigraphically) associated. Given that not all of the remains comprising a taphocoenose will be recovered, and of those that are recovered not all will be identified to taxon, what the paleozoologist can identify comprises what will here be referred to as the *identified assemblage*. It is this set of remains from which measures of taxonomic abundance are derived.

Figure 2.1 is a schematic rendition of the typical differences between a biocoenose and the identified assemblage. A biocoenose is a biological community. One or more biocoenoses is the source of input to a thanatocoenose. The transition from biocoenose to thanatocoenose involves *accumulation* (and deposition) of faunal remains in a location; accumulation can be *active* (involve a bone-accumulating agent, such as a predator that transports prey to a den) or *passive* (involve deaths of animals across the landscape, referred to as “background accumulation” [Badgley 1986]).

The transition from the thanatocoenose to taphocoenose involves both accumulation and *dispersal* and movement or removal of bones mechanically (such as by fluvial transport), and also the deterioration or chemical and mechanical breakdown of skeletal tissue. The transition from the taphocoenose to the identified assemblage involves both recovery (usually < 100 percent for any of several reasons) by the paleozoologist and the taxonomic identification of a subset of the remains comprising the taphocoenose.

The measured variable of taxonomic abundances originates in the identified assemblage; the target variable depends on the question asked (Figure 2.1). The biocoenose was Stock's target variable. Zooarchaeologists interested in human subsistence and economy often have, as their target variable, a thanatocoenose created by human predators. Paleoecologists, whether working as paleontologists or as zooarchaeologists who are interested in past ecological conditions, have as a target the biocoenose. Determination of the statistical relationship between an identified assemblage and a target variable is a taphonomic concern (Lyman 1994c), but the paleozoologist should also consider ecological and animal behavior variables as well as recovery techniques. As the Rancho La Brea materials indicate, animal behavior can influence the accumulation and thus rate of input of animal remains to the geological record.

Not all taxonomic abundances at Rancho La Brea can be attributed to historical contingencies of how and why the faunal remains present were originally accumulated. Did, perhaps, local habitats support more artiodactyls than perissodactyls, or as with the carnivores at Rancho La Brea, did a particular bone-accumulation agent bring more even-toed herbivores than odd-toed herbivores to the tar pits? Whatever the case, we measure taxonomic abundances in the identified assemblage and use those values as proxy measures or estimates of a thanatocoenose or a biocoenose. It is the task of taphonomic analyses to ascertain how good (or how biased) an estimate of a particular target variable the measured variable might be. Paleozoologists have, over the past 20 years or so become very concerned over how good an estimate of a biocoenose an identified assemblage might provide (e.g., Gilbert and Singer 1982; Ringrose 1993; Turner 1983). This concern has resulted in research known as *fidelity studies*, or assessment of "the quantitative faithfulness of the [fossil] record of morphs, age classes, species richness, species abundance, trophic structure, etc. to the original biological signals" (Behrensmeyer et al. 2000:120).

Comparisons of faunal remains with organisms making up a biological community from which the remains derive suggest that fidelity can range from quite high to rather low with respect to the variables of interest, and not all variables display equivalent fidelity in any given set of remains (Hadly 1999; Kidwell 2001, 2002; Kowalewski et al.

2003; Lyman and Lyman 2003). If there is no taphonomic reason for artiodactyl remains to outnumber perissodactyl remains at Rancho La Brea, for example, then an ecological explanation – there were more artiodactyls than perissodactyls on the landscape to be accumulated because climatic conditions created habitats more favorable to artiodactyls – is likely.

The distinction between measured taxonomic abundances and target taxonomic abundances will be important throughout the remainder of this chapter, so keep it in mind. Turner (1983) suggests that often one must assume that the relative taxonomic abundances evident in an identified assemblage are a statistically accurate reflection of those abundances in a taphocoenose, a thanatocoenose, and a biocoenose. This is true, but it is also incomplete in the sense that we can do more than assume accurate reflections; we can often test the general accuracy of the reflection with data independent of the identified assemblage at hand. The fauna represented by the identified assemblage should align ecologically with, say, floral (pollen, phytoliths, plant macrofossils) data. If it does not, then either the identified faunal assemblage is not an accurate reflection of the biocoenose, or the plant record is not. Similarly, an identified faunal assemblage at a nearby site (assuming similar ages) should align with the assemblage under consideration; two taphonomically independent assemblages should have statistically indistinguishable taxonomic abundances (and the more taphonomically independent samples that indicate the same biocoenose, the more accurate the conclusion). If they do not, one or both of the assemblages may not be a good reflection or estimate of taxonomic abundances within the local biocoenose (Grayson 1981a; Lundelius 1964).

Regardless of whether one is interested in the taphonomic history of a collection of faunal remains (how and why those remains were differentially accumulated, deposited, preserved [and some would say recovered]), in the biogeographical implications of the taxa represented by those remains (why are these taxa here but not other taxa), in the paleoecological implications of the represented taxa (do the represented species signify warm or cool climates, or moist or dry habitats), or in the subsistence and foraging behaviors of the accumulation organism (human if an archaeological site, a carnivore if a den), *taxonomic abundances* are typically part of the data scrutinized for answers to the research question. As Grayson (1979:200) noted, “It is virtually impossible to find any faunal analysis that does not present one or more measures of taxonomic abundance. [This variable] is a basic one.” The critical tactical decision concerns choosing a method to measure the abundances of the taxa represented in a collection. It is easy to show that this is no simple matter.

When Stock tallied up the remains from Rancho La Brea (Figure 1.1), it is likely that he reflected on the fact that proboscideans have more bones in one skeleton

than do perissodactyls. The former have five digits at the end of each limb whereas various perissodactyls, late Pleistocene equids in particular, have only one digit at the end of each limb. Thus, simply tallying up how many bones (and teeth) each taxon contributed to a collection could potentially produce inaccurate estimates of the abundances of the various taxa represented. If each skeleton of taxon A produces 75 identifiable bones, and each skeleton of taxon B produces 90 identifiable bones, then if 5 individuals of each were mired at Rancho La Brea, one would have 5 skulls of each taxon but 375 bones (number of identified specimens, or NISP) of taxon A and 450 bones (= NISP) of taxon B.

Ignoring for the moment potential differences in the number of teeth various taxa may have, simple tallies of the skeletal specimens of taxa A and B could produce misleading insights to which one of the two was more abundant on the landscape. Note that NISP is the measured variable whereas taxonomic abundances in the biocoenose is the target variable. Recognizing the differences between these two variables is critical to understanding whether a measure is valid or not. Is the target variable, for example, the frequency of each taxon recovered from the site (identified assemblage), the frequency of each taxon preserved in the site sediments (taphocoenose), the frequency of each taxon accumulated and deposited in the site (taphocoenose, thanatocoenose), or the frequency of each taxon available on the landscape (biocoenose)? This question underscores the simple fact that accumulation, deposition, preservation, recovery, and identification of faunal remains can all weaken in unpredictable ways the statistical relationship between the measured variable and the target variable.

If a measure of taxonomic abundances is desired, then what sort of quantitative unit should be used? Obviously, we want a unit that allows us to estimate taxonomic abundances in a sample of bones, teeth, and shell lying on the lab table. That is, we want a unit that measures taxonomic abundances within the identified assemblage; how closely those abundances match up with taxonomic abundances in the thanatocoenose or biocoenose from which the identified assemblage derives is a separate question that requires detailed taphonomic analyses and other sorts of data. We cannot presume that taxonomic abundances are the same across the identified, taphocoenose, thanatocoenose, and biocoenose assemblages. But as noted above, we can sometimes perform empirical tests to determine if this is so or not. In this chapter the two fundamental quantitative units originally designed to measure taxonomic abundances are discussed. These quantitative units are known as NISP and MNI; they are considered in turn. Biomass and meat weight and other quantitative methods used to measure taxonomic abundances are discussed in Chapter 3.

### THE NUMBER OF IDENTIFIED SPECIMENS (NISP)

The most fundamental unit by which faunal remains are tallied is the number of identified specimens, or NISP. It is just what it sounds like – the number of skeletal elements (bones and teeth) and fragments thereof – all specimens – identified as to the taxon they represent. A related measure sometimes mentioned is the number of specimens (NSP) comprising a collection or assemblage. The NSP includes bones, teeth, and fragments thereof some of which have been identified to taxon, plus those specimens that have not or cannot be identified to taxon. Typically, “identified to taxon” means identified as to the skeletal element and to the taxonomic order, family, genus, or species represented by the specimen. Most taxonomically diagnostic anatomical features are also diagnostic of skeletal element (is it a humerus or a tibia?). Many paleozoologists do not tally nondescript pieces of bone that are from the taxonomic class Mammalia if those pieces cannot be assigned to taxonomic order, family, genus, or species. This is so because taxonomic identifications such as “mammal long bone fragment” generally, but not always, are of little analytical utility. But don’t misunderstand. The research problem or question one is grappling with should, if carefully phrased, indicate whether or not an otherwise nondescript piece of “mammal long bone” is worthy of tallying or not. For the remainder of this book, “identified” means that a specimen has been minimally identified as to skeletal element and to at least taxonomic family (if not genus or species), unless otherwise noted. As Stock’s (1929) example summarized in Chapter 1 makes clear, much may be learned from taxonomic order-level identifications.

Virtually all paleozoological collections consist of some NSP of which a fraction makes up the NISP. NISP is the number of identified (to skeletal element and at least taxonomic family) specimens determined for each taxon for each assemblage. When one says NISP, what is meant is NISPi where i signifies a particular taxon. This is analogous to statistical symbolism because the i is seldom shown; rather, it is understood. Thus, one has an NISP of 10 for deer (*Odocoileus* sp.) and an NISP of 5 for rabbits (*Sylvilagus* sp.). In some cases the symbolism may be more complex, such as NISPi<sub>j</sub>, where i is for the taxon as before, and j is for a particular skeletal element or part, such as a humerus or distal tibia. Again, the ij is not shown but is understood. It is likely for these reasons of implied symbolism that it is unusual to see the plural form of NISP, such as NISPs; in my experience (which may not be representative), it is more common to read “NISP values” when more than one taxon or skeletal element is intended, as when an NISP value is given for each of multiple taxa.

### *Advantages of NISP*

The acronym NISP and its meanings, both implied (ij) and explicit (number of identified specimens), should be clear. Its operationalization may also seem to be clear and straightforward. For any pile of faunal remains, identify every specimen that you can (to skeletal element and taxon), and then tally up how many specimens you identified per taxon (and perhaps also noting how many specimens represent each kind of skeletal element, depending on your research question). NISP is an observed measure because it is a direct tally, and so it is not subject to some of the problems that derived measures such as MNI are. In part because NISP is an observed or fundamental measure, it has advantages over other units used to measure taxonomic abundances. First, NISP can be tallied as identifications are done. That is, NISP is additive or cumulative; the analyst does not have to recalculate NISP every time a new bag of faunal remains is opened and new specimens identified. This property makes NISP a fundamental measure. Every identified specimen represents a tally of “1.” Add up all the tallies of “1” for each taxon to derive the total NISP per taxon or  $\sum \text{NISP}_i$ .

NISP is, however, not free of problems. One long recognized (Clason 1972; Lawrence 1973) but seldom mentioned problem influences both NISP and MNI. Different analysts may identify different specimens in a pile of faunal remains (Gobalet 2001). The sets of specimens that any two analysts identify will be quite similar – all complete teeth and complete limb bones are likely to be identified, assuming both analysts have access to similar comparative collections – but they may not be identical, which means interanalyst comparability is imperfect. Whether a particular specimen is identifiable or not depends on the anatomical landmarks available on that specimen (Lyman 2005a), and experience and training will influence what an individual analyst will identify because that experience and training dictates which landmarks the paleozoologist has learned are useful.

Interobserver difference in what is identified can be a serious source of variation in NISP tallies. Because it concerns what is identified, interobserver difference applies to any conceivable measure of taxonomic abundance. As with other kinds of interobserver difference, it is not just difficult to control. It is impossible to control (unless every paleozoologist studies every collection) and it is likely for this reason that few paleozoologists have mentioned it. It is mentioned here for sake of completeness and because it may be an important consideration when one analyst compares his or her tallies with someone else’s for a different collection. Because it cannot be controlled, it is not considered further. But there are other potential problems with NISP. One might think that interobserver variation in how to tally what is identified for

purposes of producing an NISP value will not vary from investigator to investigator because each identified specimen represents a tally of 1. But, does it? Answering this question brings us to some of the weaknesses internal to NISP, weaknesses identified and described by many researchers.

### *Problems with NISP*

When Stock (1929) presented his census of Rancho La Brea mammals, he tallied the minimum number of individual(s) animals – what are now called MNI values – rather than NISP. It is likely he did so because he recognized that members of the Carnivora have different numbers of first (or proximal) phalanges per individual (usually 4 or 5 per limb) than do the Perissodactyla (Pleistocene horses have one) or the Artiodactyla (usually two). NISP tallies of first phalanges for a single dog would be 16 (ignoring the vestigial first + second phalanx of the first digit of each foot), for a single horse the NISP of first phalanges would be 4, and for a cervid the NISP of first phalanges would be 8.

Many problems with using NISP to measure taxonomic abundances have been described over the years by numerous authors (e.g., Bökonyi 1970; Breitburg 1991; Chaplin 1971; Gautier 1984; Grayson 1973, 1979; Higham 1968; Hudson 1990; O'Connor 2001, 2003; Payne 1972; Perkins 1973; Ringrose 1993; Shotwell 1955; Uerpmann 1973). Long lists of the possible weaknesses and potential problems with using NISP as a measure of taxonomic abundances are given by Grayson (1979, 1984). The following is based on his lists, and is supplemented with concerns expressed by paleontologists (e.g., Shotwell 1955, 1958; Van Valen 1964; Vermeij and Herbert 2004):

- 1 NISP varies intertaxonically because different taxa have different frequencies of bones and teeth (the number of elements that are identifiable varies intertaxonically);
- 2 NISP will vary with variation in fertility (number of offspring per reproductive event) and fecundity (number of reproductive events per unit of time);
- 3 NISP is affected by differential recovery or collection (large specimens [of large organisms] will be preferentially recovered relative to small specimens [generally of small organisms]);
- 4 NISP is affected by butchering patterns (different taxa are differentially butchered, one result of which is intertaxonomic differential accumulation of skeletal parts, and another of which is intertaxonomic differential fragmentation of skeletal elements);

- 5 NISP is affected by differential preservation (similar to problem 4; taphonomic influences may vary intertaxonomically);
- 6 NISP is a poor measure of diet (the bones of one elephant provide more meat than the bones of one mouse);
- 7 NISP does not contend with articulated elements (is each tooth in a mandible tallied as an individual specimen, plus the mandible itself tallied?);
- 8 the problems identified may vary between strata within a site, between distinct sites, or both, rendering statistical comparison of site or stratum specific assemblages invalid;
- 9 NISP may differentially exaggerate sample sizes across taxa;
- 10 NISP may be an ordinal scale measure and if so some powerful statistical analyses are precluded as are some kinds of inferences;
- 11 NISP suffers from the potential interdependence of skeletal remains.

### *Problems, Schmblems*

The list of problems analysts have identified as plaguing NISP values may seem disconcerting. Indeed, the length of the list may give one cause to wonder why anyone would measure taxonomic abundances using NISP in the first place. Do not, however, let such wonder convince you that NISP values are worthless. Some problems overlap with one another in terms of their effects or in terms of how they might be dealt with analytically. And notice that the list comprises a set of “possible weaknesses and potential problems.” Several of the problems are easily dealt with analytically.

Problem 1 can be controlled in several ways, such as only counting elements held in identical frequencies by the taxa under study (e.g., Plug and Sampson 1996). Do not tally phalanges of artiodactyls and perissodactyls when comparing their abundances; tally only scapulae, humeri, femora, and other elements that occur in identical frequencies in individuals of both taxa. In short, do not include tallies of elements that vary in frequency intertaxonomically. Or, weight NISP by dividing it by the number of identifiable elements per single complete skeleton in each taxon. So, if, say, horses always have 100 elements per complete skeleton and bison always have 85 elements per skeleton, then weight observed abundances of NISP for horses and for bison accordingly. This solution was suggested more than 50 years ago by paleontologist J. Arnold Shotwell (1955, 1958). It is, however, not without problems, such as requiring the assumption that complete skeletons (rather than a limb or two) were accumulated and deposited in the collection location. The assumption comprises a taphonomic problem, and might be addressed by consideration of which

skeletal elements are present. What about variation in rates of input of skeletal parts to the geological record?

We don't need to worry about correcting for differences in number of skeletal elements per taxon because, to retain the example, late-Pleistocene horses always have the same number of skeletal elements in each of their skeletons as every other late-Pleistocene horse, and late-Pleistocene bison have the same number of skeletal elements in each of their skeletons. Thus, we know that if the NISP of bison increases relative to the NISP of horses (the measured variables), then perhaps the abundance of bison (on the landscape, or in the identified assemblage) increased relative to the abundance of horses (the target variables). Bison NISP did not increase relative to horse NISP because bison evolved to have more bones or horses evolved to have fewer bones over the time represented. The same argument applies to variables that influence the rate of input of skeletal parts to the faunal record. Shotwell (1955, 1958) was concerned that different taxa input bones to the paleozoological record at different rates. A few years later Van Valen (1964) spelled out his concern that different taxa have different longevities; taxa with short individual life spans input more skeletal parts to the faunal record than taxa with long individual life spans, all else being equal (same number of skeletal parts per taxon, same population size on the landscape).

Problem 2 was recently stated by Vermeij and Herbert (2004), who worried that intertaxonomic variation in fertility and fecundity influenced the rate of skeletal part input. They noted that “short-lived (often small-bodied) species will be greatly over represented in a fossil sample relative to species with long generation times, long individual life spans, slow rates of turnover, and large body size” (Vermeij and Herbert 2004:2). If taxon A has an individual average life span of 10 years whereas taxon B has an individual average life span of 1 year, then taxon B will be represented by ten times the number of individuals as taxon A (all else being equal). Vermeij and Herbert (2004:3) were concerned that measures of “predator-prey ratios and prey availability” would be artifacts of variation in life span. Their primary solution to problem 2 requires data on average life spans, in some cases derivable from the growth increments evident in the hard parts of organisms. In the absence of the requisite ontogenetic data, a secondary solution they suggest is to restrict sampling “to organisms of comparable generation time,” though this solution also seems to require taxon-specific ontogenetic information.

The first solution is, in fact, identical in reasoning to the one suggested by Shotwell (1955, 1958) for the problem of mammalian taxa with different numbers of (identifiable) skeletal elements per individual. They are “identical” because both Shotwell and Vermeij and Herbert were concerned about biological factors that influence the

Table 2.1. *Fictional data on abundances (NISP) of three taxa in five strata*

Stratum	Taxon A	Taxon B	Taxon C	Total
V	50 (77)*	10 (15)	5 (8)	65
IV	40 (67)	10 (16.5)	10 (16.5)	60
III	30 (55)	10 (18)	15 (27)	55
II	20 (40)	10 (20)	20 (40)	50
I	10 (22)	10 (22)	25 (56)	45

\* Relative (percentage) abundances of each taxon given in parentheses.

rate at which skeletal remains are created and input to the paleozoological record. Taxa with many skeletal elements per individual and taxa with high fecundity both have higher rates of input than taxa with few skeletal elements per individual and taxa with low fecundity, respectively. When faced with the former problem, Shotwell suggested that the analyst should determine a “corrected number of specimens” per taxon, a value calculated by dividing each taxon’s NISP by the number of identifiable elements in one skeleton of that taxon. If an individual skeleton of taxon A potentially contributes 10 (identifiable) elements and taxon B 5 elements, then divide the observed NISP for A by 10 and the observed NISP for B by 5 in order to compare the abundances of the two taxa. A similar procedure for invertebrates is described by Kowalewski et al. (2003). The procedures norm all taxon-specific NISP values to a common scale – the number of identifiable skeletal elements per individual per taxon. In light of Vermeij and Herbert’s concern, a paleozoologist could norm all taxonomic abundances to a single life span, based on the duration of all life spans measured in the same unit, say, a year.

The concerns of Shotwell, Van Valen, and Vermeij and Herbert are all easily dispensed with. Table 2.1 lists fictional NISP values for three taxa in five strata. Because we know that taxon A has 10 skeletal elements per individual, taxon B has 1 skeletal element per individual, and taxon C has 5 skeletal elements per individual, we choose to weight their abundances accordingly. The results of that weighting are shown in Table 2.2. For ease of conceptualizing what has happened, consult Figure 2.2. Note that over the five strata, whether the NISP values are the raw tallies or the weighted tallies corrected for differences in number of skeletal elements per taxon, taxon A increases from stratum I to stratum V whereas taxon C decreases over that same span. Weighting does not change the results, at least with respect to increases and decreases in the *relative* abundances of taxa A and C. But, you might counter, in the unweighted data taxon B is often not very abundant at all, and it too gradually decreases from stratum I to stratum V. In the weighted data, however, taxon B is more abundant than

Table 2.2. Data in Table 2.1 adjusted as if each individual of taxon A had ten skeletal elements per individual, taxon B had one skeletal element per individual, and taxon C had five skeletal elements per individual

Stratum	Taxon A	Taxon B	Taxon C	Total
V	5 (31.5)*	10 (62.5)	1 (6.25)	16
IV	4 (25)	10 (62.5)	2 (12.5)	16
III	3 (18.75)	10 (62.5)	3 (18.75)	16
II	2 (12.5)	10 (62.5)	4 (25)	16
I	1 (6.25)	10 (62.5)	5 (31.25)	16

\* Relative (percentage) abundance in parentheses.

A and C combined, and taxon B doesn't change in relative abundance over the stratigraphic sequence. That is a good point – it suggests the data are ordinal scale – and we will return to it. First, however, we need to consider other problems with NISP.

Problem 3 concerns collection bias. Correction factors might be designed to account for the fact that small bones and teeth and shells tend to come from small organisms, and these tend to escape visual detection and to fall through coarse-meshed hardware cloth meant to allow the passage of sediment but not faunal

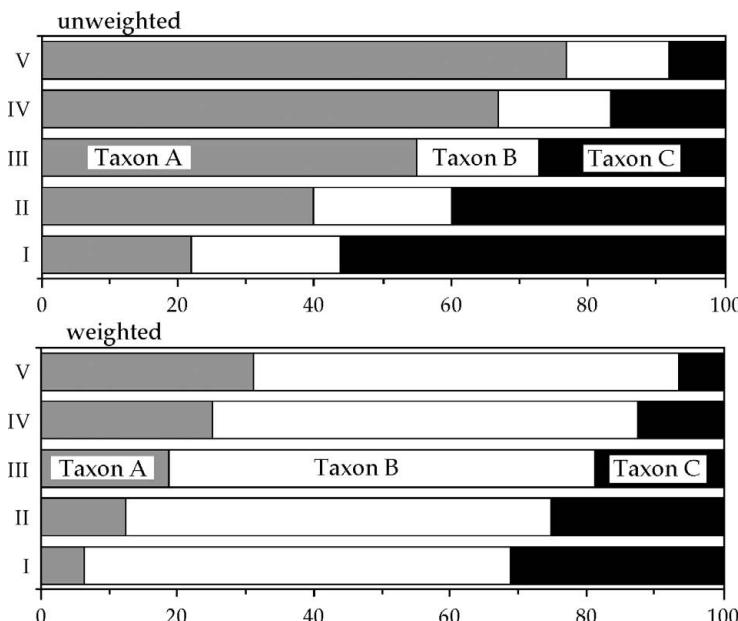


FIGURE 2.2. Taxonomic relative abundances across five strata. Data from Tables 2.1 (unweighted) and 2.2 (weighted).

remains. The design of correction values has also been a long-standing interest among zooarchaeologists (e.g., Payne 1972; Thomas 1969), but again, there are problems with these values. For example, if one uses correction values, one must assume that the samples used to derive those values are on average representative of all situations – within any given excavation unit, within any given stratum of a site, and within any given site – where small remains may fall through screens (see Chapter 4). As long as recovery methods do not differ between strata, such as using 1/4-inch mesh hardware cloth for every other stratum and using 1/8-inch mesh hardware cloth for the other strata, remains of mice will be as consistently recovered in all recovery contexts as are remains of rabbits and deer.

The preceding does not allow for differential fragmentation across taxa, the issue raised by problem 4. If rabbit bones are quite fragmented and small, they may fall through screens much more readily than unfractured remains of mice (Cannon 1999). Large bones may be more likely to be fractured by humans or carnivores because they contain more nutrients (marrow) than small bones. Fragmentation reduces identifiability by disassociating if not destroying the distinctive landmarks used to tell that bone A is a rabbit tibia whereas bone B is a duck humerus (Lyman 2005a; Marshall and Pilgram 1993). Problem 5 concerns intertaxonomic differences in preservation and may be related to problem 4 given that fragmentation influences preservation by effectively destroying bones through the process of rendering them unidentifiable. If preservational processes vary intertaxonomically, then NISP values will be differentially influenced across taxa. The magnitude of the fragmentation problem can be evaluated analytically (see Chapter 6). Fragmentation and preservation are taphonomic processes and may well render NISP data nominal scale with respect to taxonomic abundances.

Problem 6 is a serious concern to many zooarchaeologists because it is true that NISP is a poor measure of diet because the meat from the bones of one elephant will feed more people than the meat from the bones of one mouse. Furthermore, ethnoarchaeological data suggest that we cannot assume that each individual animal carcass was consumed entirely (e.g., Binford 1978; Gifford-Gonzalez 1989; Lyman 1979). But if we are concerned with dietary issues, we have in fact changed the target variable from a measure of taxonomic abundances – are there more elephants represented in the collection, or more mice – to one concerning how much of each taxon was eaten. Because we are asking a different question, a quantitative unit or measured variable different than one that simply tallies taxonomic abundances would seem to be required (see Chapter 3).

Problem 7 is that NISP does not include inherent rules for dealing with articulated elements, and while rules have been suggested (e.g., Clason 1972), these are not agreed upon by all paleozoologists or used consistently. A common example concerns

what to do with a mandible or maxilla that contains teeth. The mandible (dentary) is a discrete anatomical organ, as is each tooth. In ungulates that means a single mandible (say, the left side) containing all teeth will have 4 incisiforms (3 incisors and a canine that has evolved into the form of an incisor), 3 deciduous premolars, 3 permanent premolars, and 3 molars. It is rare to find mandibles with all 6 premolars (the deciduous ones nearly worn to nothing and about to fall out of the mandible; the permanent ones still forming and without developed roots but beginning to erupt). So, ignoring that possibility for the moment, when a mandible with all teeth is found, do we tally an NISP of 1, or do we tally an NISP of 11 (mandible + 4 incisiforms + 3 premolars + 3 molars)? How do we tally an articulated hind limb of an artiodactyl; as 1, or as 15 (femur + patella + tibia + distal fibula [lateral malleolus] + calcaneus + astragalus + naviculo cuboid + 4th tarsal + metatarsal + 6 phalanges [not to mention sesamoids])?

The paleozoologist should be consistent in applying across all taxa the tallying method chosen, and also be explicit about which method is used – tally articulated specimens as 1, or tally each distinct anatomical element, articulated or not, as 1. Perhaps the most important aspect of dealing with this problem is granting flexibility to meet the needs of analysis and the nature of the collection. Each mandible, for example, can be tallied as 1 regardless of whether it includes teeth or not (noting which teeth if any are present for purposes other than estimating taxonomic abundances, although ontogenetic age differences indicated by the teeth may play a role in estimating abundances [see the discussion of MNI]). But which skeletal elements were articulated when found and which were isolated or not articulated with other specimens is seldom noted by excavators. Thus, the paleozoologist may be forced to tally each specimen individually as 1, noting when one specimen “articulates” with another if they come from the same excavation unit, which is not necessarily the same as saying that it was “articulated” when it was found.

Noting that the problems listed thus far may vary not only intertaxonomically, but intrataxonomically within and between strata (problem 8) sounds hopelessly fatalistic. However, it is largely an analytical matter to determine if indeed this taphonomic problem applies in any given case. Even if it does, it may not preclude statistical comparisons of assemblages of faunal remains. Before arguments and examples of why this is so are presented, we consider what is likely the most serious problem with NISP. As a preface, note that most of the problems with NISP discussed so far are *not* fatal to it as a measure of taxonomic abundance. Most of these problems were identified and subsequently reiterated time and time again not as reasons to abandon NISP and design a new quantitative unit but rather were presented as warrants to use MNI (Grayson 1979, 1984). A prime example of this is problem 6. That problem – that NISP doesn’t give a good estimate of the amount of meat provided – is like

*Table 2.3. The differential exaggeration of sample sizes by NISP*

	Taxon A	Taxon B	Taxon C
NISP	1	10	10
MNI	1	1	10

saying that because a tape measure doesn't measure color, the tape measure is flawed. Of course, the tape measure was not designed to measure color, or weight, or material type; rather, it was designed to provide measures of linear distance. Because a measurement unit doesn't measure a particular variable is no reason to discard that unit completely. No one has demonstrated that NISP doesn't provide valid and accurate measures of taxonomic abundances in a taphocoenose, in a thanatocoenose, or in a biocoenose. Indeed, virtually all of the problems with NISP do not universally invalidate it as a unit with which to measure taxonomic abundances.

None of the preceding is meant to imply that NISP is a valid measure of taxonomic abundances in a taphocoenose, in a thanatocoenose, or in a biocoenose. It might be a valid measure of taxonomic abundances, but that remains to be determined. Before discussing how to make that determination, the most serious problem with NISP must be identified. This problem must be considered at length precisely because it is so worrisome.

#### *A Problem We Should Worry About*

That NISP may differentially exaggerate sample sizes across taxa (problem 9) is evident in the example in Table 2.3. This table illustrates that if one has an NISP of 1, then at least 1 individual (MNI) is represented; if NISP = 2, then MNI = 1 or 2; if NISP = 3, then MNI = 1, 2, or 3; and so on. Thus, were we to compare the abundances of taxa A, B, and C in Table 2.3, taxon B would be over-represented by NISP relative to taxa A and C. This is so because for taxa A and C, each individual (MNI) is represented by one specimen, so each specimen contributes an MNI of 1. But for taxon B, each individual is represented by an NISP of 10, so each specimen in effect contributes one-tenth of an individual MNI.

Problem 10 is that NISP is an ordinal-scale measure of abundance so some powerful statistical analyses and inferences are precluded. We can often say that taxon A is more abundant than taxon B, but we do not know by how much with respect to a target variable consisting of the thanatocoenose or the biocoenose. This is so even when we

can control for variation in fragmentation, variation in the number of identifiable elements per individual of different taxa, and all those other problems that afflict NISP. Notice that I said we can “often” say that one taxon is more abundant than another; I did not say that we can “always” say this. We return to this point later in this chapter.

The potential overrepresentation of some taxa by NISP is in fact a superficial concern, but it is intimately related to the deeper, more serious concern expressed in problem 11 – NISP suffers from the fact that skeletal specimens may be interdependent (Grayson 1979, 1984). The specimens of a taxon in a collection, or various subsets of those specimens, may be from the same individual animal (or each subset from a different individual). This precludes various statistical analyses and tests of taxonomic abundance data tallied as NISP that demand independent data, that is, each tally of “1” is independent of every other one. Some analysts have argued that specimen interdependence is not a serious problem. Gautier (1984), for example, based on estimates of preservation rates at various sites, notes the probability that an animal would be represented by a single specimen, by two specimens (the product of two independent probabilities represented by two specimens), by three specimens (the product of three independent probabilities), and so on. He finds that the probabilities for  $\text{NISP} > 1$  for any given individual are quite low, so Gautier (1984:240) concludes “the degree of interdependence (i.e., the fact that an animal is represented by several bones and hence counted several times) is much less than many analysts fear.”

Gautier’s (1984) estimates of preservation rates are based on a compilation of many estimates – the estimated duration of occupation of the site in years, the estimated size of the human population that occupied the site, the estimated number of animals necessary to provide sufficient food for the human occupants, the estimated degree of preservation of faunal remains, the estimated fraction of the site excavated, and the estimated rate of identification of faunal remains. As these estimates are added up, one influencing another, the final estimate of whether two specimens derive from the same animal is likely quite wide of the mark. The estimate is like a radiocarbon age of 1,000 years with an associated standard deviation of 900 years. Furthermore, Gautier’s estimates must assume that the taphonomic history of each specimen is independent of the taphonomic history of every other specimen, even when two specimens derive from the same individual animal. We know that that is false (Lyman 1994c), else we would never find articulated bones.

So, presuming that there is some degree of interdependence of specimens tallied for NISP values, what is the paleozoologist to do? One option is to accept Gautier’s arguments, and as Perkins (1973:367) suggests, “in the absence of archaeological evidence to the contrary we must assume that each [specimen] came from a different

individual.” This allows statistical manipulation of NISP data as if each tally of 1 for each taxon was indeed independent of every other tally of 1 for that taxon. But it is also likely that skeletal elements of individual animals were not accumulated and deposited completely independently of each other (Ringrose 1993). They were, after all, articulated and held together during the life of the organism. Actualistic work indicates that although complete skeletons may not accumulate as such, portions of skeletons comprising multiple elements are very often accumulated by both human and nonhuman taphonomic agents (e.g., Binford 1978, 1981; Blumenschine 1986; Domínguez-Rodrigo 1999a; Haynes 1988; Lyman 1989, 1994c). This brings us back to the question at the beginning of this paragraph: Given that there is some unknown (and largely unknowable) degree of interdependence in NISP values, what is the paleozoologist to do?

One thing we might do is assume that interdependence is randomly distributed across all taxa and all assemblages (Grayson 1979). Assuming this does not, of course, make it so. But if we could show that interdependence was distributed across taxa and assemblages in such a way as to not significantly influence measures of taxonomic abundances, then we would have an empirical warrant for using NISP to measure those abundances. So, the question shifts from: “Given the likelihood that there is some interdependence, what are we to do?” to “How are we to show that the nature and degree of interdependence does not significantly influence NISP measures of taxonomic abundance?” The answer to the new question must come after we consider the other quantitative unit that is regularly used to measure taxonomic abundances.

### THE MINIMUM NUMBER OF INDIVIDUALS (MNI)

Given the many difficulties with using NISP to measure taxonomic abundances, it is not surprising that Stock (1929) and Howard (1930) estimated abundances of mammals and birds at Rancho La Brea (Chapter 1) using a measure other than NISP. The measure they used is the *minimum number of individuals* (MNI). Prior to the middle 1990s, a plethora of acronyms were used for this quantitative unit (Lyman 1994a). As with NISP, MNI usually (but not always) is given for each identified taxon, so the acronym is more completely given as MN*i*, where *i* again signifies each distinct taxon and, again, is seldom shown but is instead understood. Recall that Stock and Howard both defined MNI as the most commonly occurring skeletal element of a taxon in an assemblage. Thus, if an assemblage consists of three left and two right scapulae of a species of mammal, then there must be at least (a minimum of) three individuals of that species represented by the five specimens because each

individual mammal has only one left and one right scapula. The number of individuals is a *minimum* because there may actually be five individuals represented by the five scapulae, but it presently is difficult to determine in each and every case which left scapula goes with which right scapula (come from the same individual), nor can we always determine when potentially paired elements do not come from the same individual. Thus, the *actual number of individuals* (ANI) represented by the identified assemblage of a taxon is difficult to determine, except perhaps in those rare cases when the taxon is represented by more or less complete articulated skeletons.

MNI is an attractive quantitative unit because it solves many of the problems that attend NISP. In particular, it solves the critical problem of specimen interdependence given how MNI is usually defined – the most commonly occurring kind of skeletal specimen of a taxon in a collection. This is indeed how many (but certainly not all) analysts define the term (Table 2.4). If the most commonly occurring kind of skeletal specimen of taxon A is distal left tibiae, then tally up distal left tibiae; the total equals the MNI of taxon A. If the most commonly occurring kind of skeletal specimen of taxon B is the right m<sub>3</sub>, then tally those up and the total gives the MNI for taxon B. No single individual of any known mammalian taxon possesses more than one distal left tibia or more than one right m<sub>3</sub>, so each one of those kinds of specimens in a collection must represent a unique individual that was alive in the past. An easy way to conceptualize MNI is this: If two skeletal specimens *overlap* anatomically, then they must be from distinct, independent individual organisms because they are redundant with one another (Lyman 1994b). If the two specimens fit together in a manner like two conjoining pieces of a jigsaw puzzle, then they are from the same individual and are interdependent. But if the two specimens do not overlap anatomically and they do not fit together like two pieces of a jigsaw puzzle, then they *may* be from the same individual unless they are clearly of different size, ontogenetic (developmental) age, or sex. If they are of the same size, age, and sex, but do not overlap or conjoin, then they *may* or *may not* be from the same individual. That is a sticky point to which we will return in force in Chapter 3.

MNI seems to have originated in paleontology with individuals such as Stock (1929) and Howard (1930). It has been suggested that MNI was introduced to (zoo)archaeologists in 1953 by Theodore White (Grayson 1979), a paleontologist who worked with zooarchaeological collections, and this could well be correct. However, an archaeologist working a few years prior to White used MNI as a measure of taxonomic abundances.

In his unpublished Master's thesis, William Adams (1949:23–24) estimated the “approximate number of animals represented by the sample” of bones and teeth he

Table 2.4. *Some published definitions of MNI*

- 
1. “the number of similar parts of the internal skeleton as for example the skull, right ramus of mandible, left tibia, right scaphoid” (Stock 1929:282).
  2. “for each species, the left or the right of the [skeletal] element occurring in greatest abundance” (Howard 1930:81–82).
  3. “the bone with the highest total will indicate the minimum number” (Adams 1949:24).
  4. “separate the most abundant element of the species into right and left components and use the greater number as the unit of calculation” (White 1953a:397).
  5. “the [skeletal] element present most frequently” (Shotwell 1955:330); “that number of individuals which are necessary to account for all of the skeletal elements (specimens) of a particular species found in a site” (Shotwell 1958:272).
  6. the number of lefts and of rights of each element, those matching in terms of age and size tallied as from the same individual, those not matched tallied separately as from different individuals (Chaplin 1971).
  7. “equal to the greatest number of identical bones per taxon” (Mollhagen et al. 1972:785).
  8. “a count of the most frequent diagnostic skeletal part” (Perkins 1973:368).
  9. “the most frequently occurring bone” (Uerpman 1973:311).
  10. “the number that is sufficient to account for all the bones assigned to the species; the most abundant body part” (Klein 1980:227).
  11. “the least number of carcasses that could have produced the recovered remains . . . determined by taking the raw count of the most commonly retrieved bone element that occurs only once in the skeleton” (Gilbert and Singer 1982:31–32).
  12. “may be based upon counts of the most abundant element present from one side of the body or on counts determined by joint consideration of skeletal parts represented; the size, age, and wear-state of specimens” (Badgley 1986:329).
  13. “essentially the sample frequency of the most abundant skeletal part” (Plug and Plug 1990:54).
  14. “the smallest number of individual animals needed to account for the specimens of a taxon found in a location” (Ringrose 1993:126).
  15. “the most frequently occurring element” (Rackham 1994:39).
  16. “the higher of the left- and right-side counts (if appropriate – obviously not if the most abundant element is an unpaired bone such as the atlas) is taken as the smallest number of individual animals which could account for the sample” (O’Connor 2000:59).
-

had studied. He chose “certain bones as readily identifiable, easily distinguished with regard to right or left position in the body and not commonly used for artifacts,” and tallied up the occurrences of each for two taxa in each of five distinct recovery proveniences (p. 24). He noted that “since any one animal can possess only one of each of these bones, then the bone with the highest total will indicate the minimum number of mammals represented by the bone sample from that [recovery provenience]” (p. 24). Adams summed the MNI values indicated by each assemblage of bones from a unique recovery provenience and noted that in so doing, he had to assume “that parts of one individual are not represented from more than one [recovery provenience]” (p. 24); he assumed that skeletal remains in one provenience were independent of those in all others. Despite these significant insights, Adams (1949:24) abandoned MNI values because they provided *only* “minimum numbers,” and he believed that assigning a “maximum number would be a matter of guesswork.” Adams desired a quantitative unit that provided ratio-scale taxonomic abundances. Adams did not reference Stock or any other paleontologist who had previously used MNI as a quantitative unit. Circumstantial evidence therefore suggests that Adams invented (if you will) MNI independently of its invention in paleontology. But because he did not publish his discussion, few zooarchaeologists seem to have been aware of the MNI quantitative unit prior to White’s work. At least few of them used MNI prior to the late 1950s, by which time it had been used in clever ways by Theodore White, who published his results in archaeological venues.

Unlike Adams, White (1953a, 1953b) did not seek to estimate taxonomic abundances when he introduced MNI to zooarchaeologists. Rather, he sought to estimate the amount of meat provided by each taxon; that was his target variable. He (1953a:397) noted that “four deer [*Odocoileus* sp.]” were needed to provide as much meat as “one bison [*Bison bison*] cow,” and NISP would not reveal how much meat was provided by each of these taxa. Being a paleontologist who likely was familiar with, and used to seeing MNI values reported in the paleontological literature (e.g., Howard 1930; Stock 1929), White was aware of a quantitative unit (MNI) that could be easily converted to meat weight. White may have believed that there was no other reason that animal remains would be of interest to archaeologists, other than to reveal some aspects of human behavior. Diet – what folks ate – was an obvious human behavior reflected by animal remains.

Methods, including White’s, to estimate meat weight (and the related variable, biomass) are discussed in Chapter 3. The important point here is that MNI was introduced to zooarchaeologists not as a replacement for NISP as a measure of taxonomic abundances. Rather, MNI was introduced to zooarchaeology in order to measure something else, specifically the amount of meat represented by a collection

of faunal remains. From a historical perspective, this is interesting for the simple reason that MNI was used in yet another discipline originally to estimate taxonomic abundances. The measurement of biomass and meat weight was introduced in that discipline as a replacement for MNI as a measure of diet, that is, for virtually the same reason MNI was introduced in zooarchaeology.

One of the things that ornithologists are interested in is the diet of birds. Raptors (hawks and eagles) and owls hunt, among other animals, small mammals – various insectivores, rodents, and leporids – and, depending on the taxon of the bird, swallow partial or complete carcasses of their prey. After 12–24 hours or so, a “pellet” of hair, bones, and teeth is regurgitated (the common terms in the ornithological literature are “egested” or “cast”). Depending on the bird, the bones and teeth are often in very good condition (not broken or excessively corroded from digestive acids) and retain many taxonomically diagnostic features (Andrews 1990). Because pellets are deposited beneath roosts (resting areas) and nests, a collection of pellets from such a location can reveal much about the diet of the bird.

Studies of such pellets and the faunal remains they contain have a deep history in ornithology (e.g., Errington 1930; Fisher 1896; Marti 1987; Pearson and Pearson 1947). Because ornithologists study the remains of prey in those pellets in order to answer some of the same questions that paleozoologists do (Which taxon is most abundant and which is least abundant on the landscape? Which taxon provided the most sustenance to the predator? [e.g., Andrews 1990; Mayhew 1977]), ornithologists have grappled with some of the same issues that paleozoologists have, especially with respect to how to quantify the remains of vertebrate prey. Ornithologists quickly figured out that NISP might not give a valid indication of which prey taxon was the most frequently consumed, so they did one of two things. They either tallied only skulls, or they determined the MNI based on whether the skull, left mandible, or right mandible was the most common skeletal element in a collection. They used both of these approaches as early as the 1940s (references in Lyman et al. 2003), describing how they counted taxonomic abundances. The earliest formal definition of MNI by an ornithologist of which I am aware is Mollhagen et al.’s (1972:785): the “minimum number of animals [is] equal to the greatest number of identical bones per taxon.” No ornithologist who uses MNI, references Stock or any other paleontologist who used MNI, suggesting yet another independent invention of MNI. Near universal adoption of MNI as the quantitative unit of choice of ornithologists lead quickly to recognition that an MNI of five for each of two taxa did not give an accurate measure of diet when individuals of those two taxa were of rather different size. Thus, ornithologists determined the live weights of average adult individuals of common prey species and

used those data to determine the composition of a bird's diet (Steenhof 1983), much as White (1953a, 1953b) had done 30 years earlier.

Given that three separate disciplines have used MNI, and all of them (granting Adams's flirtation with it) seem to have independently invented it, one might think that MNI is a well-understood unit of measurement. It is commonsensical to calculate, and it has a basis in the empirically verifiable reality of the individuality and physical discreteness and boundedness of every organism. But MNI is not a well-understood quantitative unit. It has a number of problems, just like NISP. And also just like NISP, several of the problems with MNI are trivial or easily dealt with analytically, but one of them is rather serious.

### *Strengths(?) of MNI*

Klein (1980:227) stated that unlike NISP, MNI is not affected by differential fragmentation, and suggested that this was a reason to seriously consider using MNI values as measures of taxonomic abundance, particularly when comparing assemblages with different degrees of fragmentation. He was concerned that a taxon, the remains of which had not been broken, would be underrepresented by NISP relative to a taxon the remains of which had been broken, all else being equal. Although Klein is correct that fragmentation will increase NISP, he is only partially correct because in reality fragmentation can influence MNI in two ways. First, fragmentation of moderate intensity, say, breaking each element into two more or less equal size pieces, will not influence MNI because specimens will retain anatomically and taxonomically diagnostic features (Lyman 1994b). Second, as the intensity of fragmentation increases, meaning that as fragments get smaller and represent less of the skeletal element from which they originate, the more difficult it will be to identify those fragments as to skeletal element represented and to taxon. This is so because progressively smaller fragments are successively less likely to retain anatomically and taxonomically diagnostic features (Lyman and O'Brien 1987). Thus fragmentation first increases NISP (but not MNI), but then as fragmentation intensifies, NISP decreases and so too does MNI.

The relationship between fragmentation and NISP, and that between fragmentation and MNI, was spelled out by Marshall and Pilgrim (1993) with respect to measuring the frequencies of individual skeletal parts. Because MNI is based on the most frequent skeletal part, Marshall and Pilgrim's findings are equally applicable to both NISP and MNI. Fragmentation influences MNI, although in a manner different than it does NISP. Breaking a skeletal element into pieces will first increase,

and then decrease NISP as fragmentation intensity increases and specimens become unidentifiable; moderate breakage will not influence MNI, but intensive fragmentation will result in a decrease in MNI as progressively more specimens fail to retain sufficient anatomical landmarks to allow identification.

Klein (1980) is correct that MNI will not be influenced by fragmentation, whereas NISP will be, but only in the limited case of assemblages with little fragmentation. Both NISP and MNI values will be influenced by *intensive* fragmentation. We do not know, however, the degree of fragmentation intensity at which fragmentation begins to decrease identifiability and to reduce values of both NISP and MNI (Lyman 1994b). The difficulty of establishing this degree of fragmentation is exacerbated by the fact that small fragments are unidentifiable to skeletal element.

MNI overcomes such problems as intertaxonomic variation in the number of identifiable elements per individual. But as I noted, in regard to this problem as it pertains to NISP, it is easy to analytically correct for intertaxonomic variation in the number of identifiable skeletal elements per individual. Either normalize all NISP by dividing those values by a value that accounts for variation in identifiable elements per skeleton, or delete from tallies those taxonomically unique elements (such as upper incisors in equids when comparing their abundance to bovids, who don't have upper incisors).

The most important advantage of MNI is that it overcomes possible specimen interdependence because of how MNI is defined (Table 2.4). As Ringrose (1993:127) noted, the "basic principle of the MNI is to avoid 'counting the same animal twice.'" If MNI is derived according to the definitions in Table 2.4, there is no way for two or more specimens in one assemblage to be from the same individual. This is illustrated in Table 2.5. Notice in that table of fictional data that there is a minimum of seven individuals (= MNI). Notice also that all postcranial elements have been assigned to an individual already represented by a skull and both mandibles. The left scapulae do not represent, so far as we can tell in light of modern methods, an eighth individual, a ninth individual, and so on, nor do the right scapulae, the left humeri, and so on. Here,  $\sum \text{NISP} = 71$ , but obviously if we tally NISP, we count the first individual fifteen times ( $= \sum \text{NISP}$  for that individual), the second individual is counted fourteen times, and so on. Of course, the right radius assigned to individual number five may actually go with individual number four, or with individual number eight, but we cannot determine that. Thus, the number of individuals represented by the seventy-one specimens listed in Table 2.5 is a minimum of seven individuals, but the specimens might represent more.

MNI would seem to avoid the interdependence problem *within an assemblage* such as shown in Table 2.5. But note the emphasized phrase – *within an assemblage*.

Table 2.5. A fictional sample of seventy-one skeletal elements representing a minimum of seven individuals (I)

	I.1	I.2	I.3	I.4	I.5	I.6	I.7
Skull	+	+	+	+	+	+	+
L mandible	+	+	+	+	+	+	+
R mandible	+	+	+	+	+	+	+
L scapula	+	+	+	+	+	+	
R scapula	+	+	+	+	+	+	
L humerus	+	+	+	+	+	+	
R humerus	+	+	+	+	+		
L radius	+	+	+	+	+		
R radius	+	+	+	+	+		
L innominate	+	+	+	+			
R innominate	+	+	+	+			
L femur	+	+	+				
R femur	+	+	+				
L tibia	+	+					
R tibia	+						
$\sum \text{NISP}$	15	14	13	11	9	6	3

\* + denotes a skeletal element is represented.

What about *between* assemblages? What if the left femur of an individual is in one assemblage and the matching right femur is in another assemblage? Were we to tally each as part of an MNI calculation in the two assemblages, we would have counted that single individual twice. This introduces the most serious problem with MNI – aggregation – and there are other, less serious but significant problems as well.

### Problems with MNI

As with NISP, analysts have identified what they take to be serious problems with MNI (e.g., Casteel n.d.; Fieller and Turner 1982; Gilbert et al. 1981; Grayson 1973, 1979, 1984; Klein 1980; Klein and Cruz-Uribe 1984; Plug and Plug 1990; Ringrose 1993; Turner 1983, 1984; Turner and Fieller 1985). These include:

- 1 MNI is difficult to calculate because it is not simply additive;
- 2 MNI can be derived using different methods, thereby reducing comparability;
- 3 MNI values do not accurately reflect the thanatocoenose or the biocoenose;

- 4 MNI values exaggerate the importance of rarely represented taxa, or taxa represented by low NISP values;
- 5 MNI values are minimums and thus ratios of taxonomic abundances cannot be calculated;
- 6 MNI is a function of sample size or NISP, such that as NISP increases, so too does MNI; and
- 7 different aggregates of specimens comprising a total collection will produce different MNI values.

The first problem – MNI is difficult to determine – is not worth considering. No one has ever said research of any kind was, or should be, easy. But related to this problem is the fact that MNI is not additive like NISP is. Rather, every time a new bag of faunal remains is opened and specimens identified, one has to rederive the MNI (assuming it was derived before). This is so because the most common skeletal part per taxon may change with the addition of another bag of bones. This problem, too, is trivial given that research often involves calculation and recalculation, again and again, as new data are collected or as new insights are gained and adjustments are made to a data set.

That different researchers use different methods to derive MNI values (the second problem) is evidenced by variations in the definitions of MNI in Table 2.4. This problem is akin to the one that different analysts will produce different NISP values for the same collection of remains given their varied expertise at identification. There is no real way to control this problem, so the methods used to derive MNI values should be stated explicitly. Were remains matched by size? By age? By recovery context? All of these or other variables? Potential and varied use of these criteria make MNI a *derived* measure.

Some have argued that to not attempt to match potentially paired remains such as left femora with right femora – to determine if they originated in the same individual – results in a misleading MNI value (Fieller and Turner 1982). White (1953a:397) thought that matching would require “the expenditure of a great deal of effort with small return [to] be sure all of the lefts match all of the rights.” That is, he believed that checking every possible bilateral pair of bones for matches (based on the notion of bilateral symmetry – that a left element is a mirror image of its right element mate), and assuming that each of those matches derived from the same individual – would result in a relatively small increase in the MNI value, and thus not much in the way of alteration of taxonomic abundances. Whether or not White was correct with respect to the magnitude of change in MNI values when matching is undertaken is unclear, but likely is assemblage specific for many reasons (Lyman 2006a). Some argue that

the difference would be considerable whereas others seem to think it would not be if “a great deal of effort” were expended, whatever the result, it will be a function of the identified assemblage at hand and the procedures used to analytically manipulate the bilateral pair data rather than the act of matching and identifying bilateral pairs itself (compare Fieller and Turner [1982] with Horton [1984]).

The third problem, too, can be said to characterize both NISP and MNI. NISP is a count of identifiable specimens in the assemblage rather than a measure of the thanatocoenose or the biocoenose. Similarly, given its definition, MNI is the *minimum* number of animals necessary to produce the identified specimens comprising the identified assemblage. Both quantitative units *describe* the assemblage using two rather different variables. Whether either NISP or MNI more accurately reflects the thanatocoenose or the biocoenose cannot be assumed given the contingent and particularistic nature of the taphonomic history of the assemblage (e.g., Gilbert and Singer 1982; Ringrose 1993; Turner 1983). But, as I noted with respect to NISP, there are ways to test if the identified assemblage rendered as a set of MNI values accurately reflects the biocoenose. Are the taxonomic abundances indicated by MNI values what would be expected given independent evidence of environmental conditions? Do MNI abundances match those from contemporaneous nearby faunal assemblages that experienced independent taphonomic histories? If the answers to these questions are all “yes,” then it would be reasonable to suppose that the taxonomic abundances in the collection under study are fairly accurate reflections of those abundances in the thanatocoenose as well as the biocoenose.

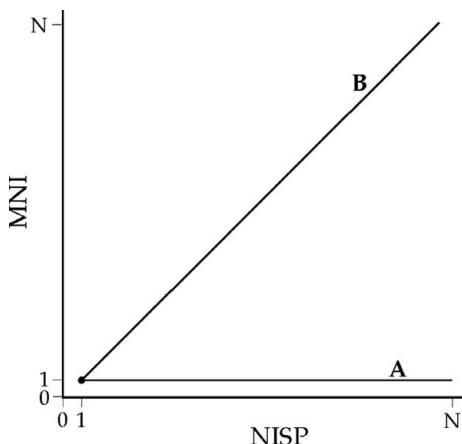
The fourth problem can be appreciated by recalling Table 2.3. There, taxon A is rarely represented ( $\text{NISP} = 1$ ) whereas taxon B is frequently represented ( $\text{NISP} = 10$ ), but both taxa have an MNI of 1. MNI exaggerates the representation of taxon A relative to taxon B’s representation. Although this observation is true, it is merely the converse of the related problem with NISP. Thus one might argue that MNI and NISP are equally flawed in this respect. If you are uncomfortable with that, you could note that taxon A in Table 2.3 is represented by one tibia, and tally only tibiae identified as taxon B when comparing abundances of these two taxa. It is likely that such a procedure would decrease the disparity in representation of the two taxa, but it also demands the assumption that the other specimens of taxon B are interdependent with that taxon’s tibiae, an assumption that would likely be difficult to warrant empirically or theoretically. (Some specimens may be interdependent, but it seems improbable that all would be interdependent.)

Plug and Plug (1990) identified the fifth problem: MNI values are minimums and thus ratios of MNI values cannot be validly calculated (see also Gilbert et al. 1981). They note that if the MNI of taxon *A* is 10 and the MNI of taxon *B* is 20, we cannot

use simple arithmetic to calculate the  $A:B$  ratio because, with respect to the true number of individuals, it is very likely that  $A \geq 10$  and  $B \geq 20$ . Thus any ratio  $A:B$  cannot be validly calculated. MNI values are not ratio scale. Instead, they are *perhaps* ordinal scale. Thus we can say, in Plug and Plug's case of  $A:B$ , that it is likely that  $A < B$ , but we cannot say by how much given that both  $A$  and  $B$  are minimum values, and we don't know their true values. A similar argument can be made with respect to NISP values. They are maximum estimates of taxonomic abundances, so a ratio of NISP values for two taxa, although easily calculated, may not actually be a ratio scale measure of taxonomic abundances. That MNI values are minimums is clear from how they are derived (Table 2.4). Paleozoologists have known these things since the MNI quantitative unit was introduced (e.g., Adams 1949). What is perhaps less well-known is that recent simulations indicate that MNI often provides values considerably lower than the actual number of individuals (ANI) present in a collection (Rogers 2000a). To illustrate this, look at Table 2.5 one more time. Here, the MNI is seven; the NISP is seventy-one. If each skeletal element is independent of every other element (each comes from a different organism), then the ANI is seventy-one, an order of magnitude greater than the MNI value.

The fact that MNI increases as NISP increases (problem 6) has been recognized for some time in zooarchaeology (Casteel 1977, n.d.; Ducos 1968; Grayson 1978a). Some have argued that this statistical relationship warrants use of NISP rather than MNI to measure taxonomic abundances; the reason is that the same information regarding taxonomic abundances is contained in both quantitative units, so there is no reason to determine MNI. Others have noted that although this statistical relationship does indeed exist between the two measures, the precise nature of the relationship depends on the particular set of remains involved (e.g., Bobrowsky 1982; Grayson 1984; Hesse 1982; Klein and Cruz-Uribe 1984). Some researchers use the last observation – that the relationship between NISP and MNI is statistically particularistic – to argue that one cannot predict MNI from NISP in a new sample based on the statistical relationship of the two in previously studied samples, so perhaps MNI should be determined (Klein and Cruz-Uribe 1984). This is a clever insight, and it is correct, but it does not mean we must determine MNI values when seeking measures of taxonomic abundance. We need not determine MNI because of the relationship between the NISP for a taxon in an assemblage and its attendant MNI.

It is commonsensical that as the NISP of a taxon increases so too should the MNI for that taxon. This is so because every individual skeleton comprises a limited number of elements (or what might become identifiable specimens comprising the paleozoological record). Adding randomly chosen skeletal elements selected from, say, 100 skeletons of an identified assemblage, the first element will contribute one individual.



**FIGURE 2.3.** The theoretical limits of the relationship between NISP and MNI. Modified from Grayson (1978a). Line A indicates that every new specimen does not contribute a new individual. Line B indicates that every new specimen contributes a new individual.

That is,  $\text{NISP} = 1 = \text{MNI}$ . The second element will contribute another NISP ( $\sum \text{NISP} = 2$ ), but that element might, or might not, contribute another individual ( $\sum \text{MNI} = 1$  or 2). The third element will contribute yet another NISP ( $\sum \text{NISP} = 3$ ), and it might contribute another individual or it might not ( $\sum \text{MNI} = 1, 2$ , or 3). And so on until the probability of adding a bone or tooth of an already represented skeleton is greater than the probability of adding a bone of an unrepresented skeleton, at which point the rate of increase in MNI will slow relative to the rate of increase of NISP. As Grayson (1978a) noted, there are two limits to the possible relationship between NISP and MNI. Either every new skeletal element derives from the same individual and thus every  $\text{NISP} > 1$  contributes nothing to the MNI tally, or every new skeletal element derives from a different, unique individual and thus every  $\text{NISP} \geq 1$  contributes another MNI. These relationships express the limits of all possible relationships between NISP and MNI (Figure 2.3).

The individual limits to the relationship between NISP and MNI (Figure 2.3) are unlikely to be found in the real world. Unless one is dealing with, say, the moderately fragmented skeleton of a single individual animal (NISP is several hundred), it is likely that NISP will increase more rapidly than MNI. In fact, unless one is dealing with an assemblage of remains of a single taxon that has but one identifiable skeletal element (such as the unbroken shells of a gastropod), it is likely that new NISP will often be added without adding any MNI. The general relationship between NISP and MNI is described by the line in Figure 2.4. It is relatively easy to show that this is indeed the relationship that is found in case after case. The relationship

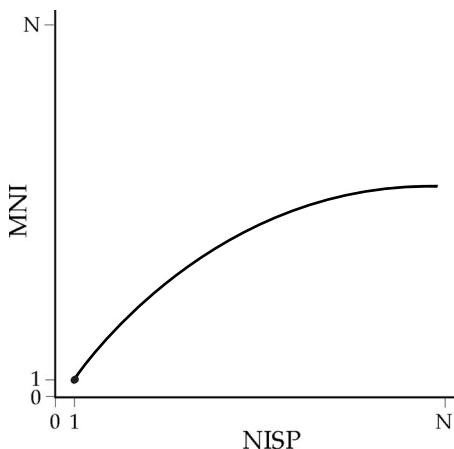


FIGURE 2.4. The theoretically expected relationship between NISP and MNI. Modified from Casteel (1977). As NISP increases, it takes progressively more specimens to add new individuals.

is curvilinear because, given a finite NISP and a finite MNI, specimens from an individual already represented are progressively more likely to be added as the sample size ( $\sum$ NISP) increases. Whatever the kind of skeletal part that is the most frequent or most common and thus defines MNI, that kind of part will become progressively more difficult to find (Grayson 1984).

Ducos (1968) found that this curvilinear relationship could be made linear (and thus perhaps more easily understood when graphed) if both the NISP data and the MNI data were log transformed (see Box for additional discussion). As Grayson (1984:52) later noted, untransformed NISP data and untransformed MNI data may sometimes be related in a linear fashion, but they often are not. The means to tell if they are not involves examination (either visual or statistical) of the residuals (the distance above and below the regression line of the plotted points and the pattern of the distribution of those points). Typically, log transformation reduces the dispersal of points to a statistically insignificant level. The slope of the best-fit regression line summarizes the rate of change in MNI relative to the rate of change in NISP and is described by a single number representing a power function (or exponent); the larger the number, the steeper the slope.

Casteel (1977) found the relationship modeled in Figure 2.4 in a series of assemblages of zooarchaeological and paleontological materials representing numerous taxa. His data originally were comprised of 610 paired NISP–MNI values. (A paired NISP–MNI value is the NISP value and the MNI value for a taxon in an assemblage of remains.) Casteel subsequently expanded his data set (Casteel n.d.) to include 3,440

**BOX 2.1**

It is often easier to grasp intuitively a linear relationship between two variables than a curvilinear one. In many cases log transformation of NISP and MNI data causes what is otherwise a curvilinear relationship to become linear. The typical form of a linear relationship can be expressed by the equation  $Y = a + bX$ , where  $X$  is the independent variable (in this case, NISP),  $Y$  is the dependent variable (in this case, MNI),  $a$  is the  $Y$  intercept (where the line describing the linear relationship intersects the  $Y$  axis), and  $b$  is the slope of the line (where the slope of the line describing the linear relationship represents how fast  $Y$  changes relative to change in  $X$ ). The simple best-fit regression line in a graph showing the relationship between log transformed NISP data and log transformed MNI data is described by the formula  $Y = aX^b$ , where the variables  $Y$ ,  $a$ ,  $X$ , and  $b$  are as defined above. This formula describes what is referred to as a power curve; if  $b$  is positive the curve extends upward from the lower left to the upper right of the graph; if  $b$  is negative the curve extends downward from the upper left to the lower right. If we transform both sides of  $Y = aX^b$  to logarithms, then we have  $\log Y = \log a + b \log X$ , linear relationship between  $\log Y$  and  $\log X$ . In this volume, I present the relationship between  $\log X$  or  $\log$  NISP, and  $\log Y$  or  $\log$  MNI, in the form  $Y = aX^b$ , or what is simply a different form of the linear relationship. The  $Y$  intercept should be zero, given that a zero value for NISP must produce a zero value for MNI, but practice has been to allow the empirical data to identify a  $Y$  intercept; I follow this practice here noting that should the empirically determined  $Y$  intercept differ considerably from zero, the data used should be inspected to determine why. Variables  $a$  and  $b$  are constants determined empirically for each data set.

paired NISP–MNI values. (The manuscript in which Casteel used this larger data set was never published. It was written in 1977, and afterwards cited occasionally by his colleagues [e.g., Bobrowsky 1982; Grayson 1979]. I obtained a copy of the manuscript from Grayson in the late 1970s.) In both cases, Casteel found a statistically significant relationship between NISP and MNI like that shown in Figure 2.4. Bobrowsky (1982) found the same relationship between paired NISP–MNI values using much smaller data sets. Both Casteel (1977, n.d.) and Bobrowsky (1982) graphed the relationship using untransformed data; Grayson (1984) and Hesse (1982) used log-transformed data in their graphs of the relationship. Grayson (1984) summarized many cases that had been reported by others, and reported several new cases to show that the relationship was essentially ubiquitous. Klein and Cruz-Uribe (1984) found the same

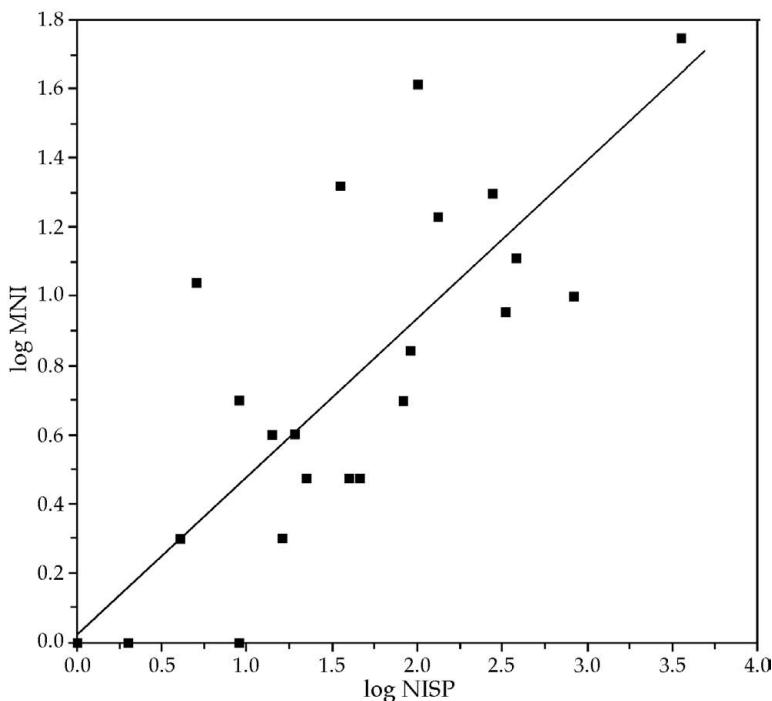


FIGURE 2.5. Relationship between NISP and MNI data pairs for mammal remains from the Meier site (data in Table 1.3). See Table 2.6 for statistical summaries.

relationship between NISP and MNI using data sets different from those used by Casteel, Bobrowsky, Grayson, and Hesse.

The large collection from Meier ( $\sum \text{NISP} = 5939$ ) shows the relationship between NISP and MNI nicely (Figure 2.5). The NISP–MNI data pairs are strongly correlated (Pearson's  $r = 0.8734$ ,  $p < 0.0001$ ). The slope of the best-fit regression line ( $= 0.487$ ) is similar to that reported by others for other data sets; Casteel (1977) reported a slope of 0.52, for example, and Grayson (1984) reported six others that ranged from 0.40 to 0.64. The precontact assemblage from Cathlapotle (Table 1.3) also shows the nature of the relationship between NISP–MNI data pairs (Figure 2.6), as does the postcontact assemblage from that site (Figure 2.7). In all three cases, the correlation coefficient is strong ( $r > 0.87$ ) and significant ( $p < 0.0001$ ). The statistical relationships between the two variables in each of these three assemblages are summarized in Table 2.6.

The relationship between NISP and MNI shown in Figures 2.5, 2.6, and 2.7 is not unique to the Portland Basin. Recall that Casteel, Grayson, Hesse, Bobrowsky, and Klein and Cruz-Uribe found exactly the same relationship between the two variables in collections from all over the world and representing many time periods and taxa.

Table 2.6. Statistical summary of the relationship between NISP and MNI for mammal assemblages from Meier (Figure 2.5) and Cathlapotle (Figures 2.6 and 2.7) (see Table 1.3 for data).

Site	Regression equation	<i>r</i>	<i>p</i>	$\sum \text{NISP}$	N of taxa
Meier	$\text{MNI} = -0.06(\text{NISP})^{0.487}$	0.873	<0.0001	5,939	26
Cathlapotle, precontact	$\text{MNI} = -0.098(\text{NISP})^{0.42}$	0.916	<0.0001	2,372	21
Cathlapotle, postcontact	$\text{MNI} = -0.0557(\text{NISP})^{0.44}$	0.901	<0.0001	3,834	24

Together, these cases suggest that the relationship is nearly ubiquitous. Table 2.7 summarizes the statistical relationship between NISP and MNI in fourteen assemblages of mammal remains from fourteen sites in eastern Washington State. I, along with two fellow graduate students at the time, identified the taxa in these collections in the late 1970s. All fourteen collections display the same kind of relationship between NISP and MNI as is evident for Meier and Cathlapotle.

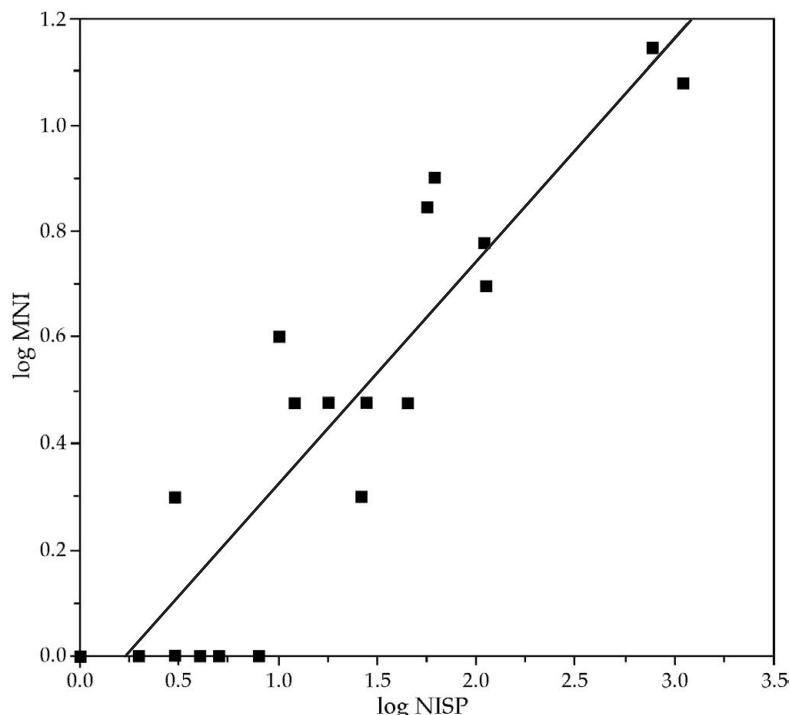


FIGURE 2.6. Relationship between NISP and MNI data pairs for the precontact mammal remains from the Cathlapotle site (data in Table 1.3). See Table 2.6 for statistical summaries.

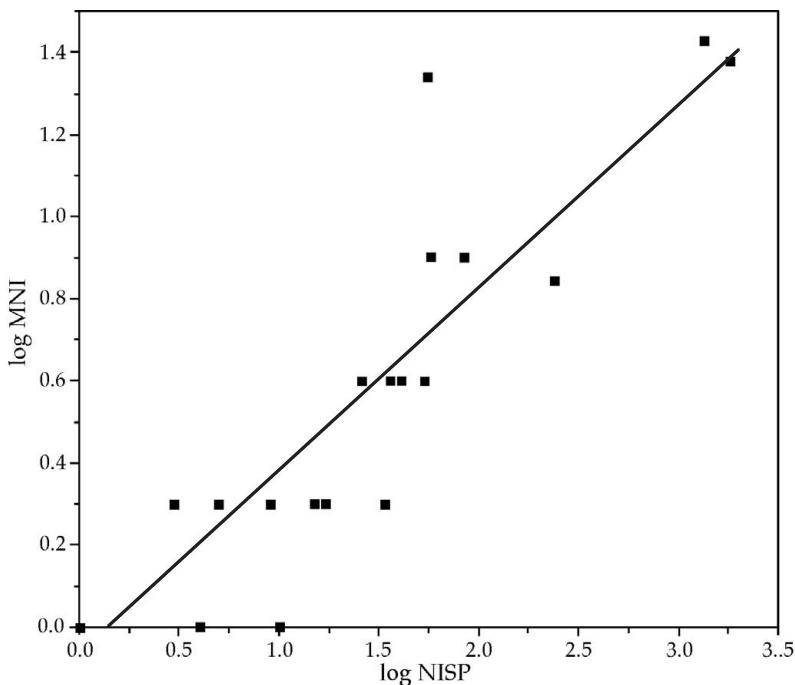


FIGURE 2.7. Relationship between NISP and MNI data pairs for the postcontact mammal remains from the Cathlapotle site (data in Table 1.3). See Table 2.6 for statistical summaries.

That NISP–MNI data pairs are often tightly related (the correlation coefficient is large) even in nonarchaeological contexts is also easy to show. Consider a sample of eighty-four pellets likely cast by a barn owl (*Tyto alba*) in eastern Washington State. NISP and MNI values for the total assemblage of prey remains in the eighty-four pellets (Table 2.8) are arrayed in a bivariate scatterplot in Figure 2.8. The relationship between the values of the two is linear and strong ( $r = 0.989$ ,  $p = 0.0002$ ), as it is among the archaeological samples discussed previously; the regression equation is:  $MNI = -1.57(NISP)^{0.935}$ . The same relationship holds for paleontological collections as well, as Grayson (1978a) showed some years ago.

The fact that NISP and MNI are often strongly correlated could be used to argue that we should use MNI as the quantitative unit for measures of taxonomic abundance because of the potential for the interdependence of specimens in NISP tallies. Indeed, Hudson (1990) argued on the basis of ethnoarchaeological data, and Breitburg (1991) on the basis of historic-era zooarchaeological data supplemented by written documents, that MNI provides more accurate measures of taxonomic abundances than NISP. That MNI would indeed sometimes be a more accurate measure of taxonomic abundances than NISP is to be expected given everything we know about the two quantitative units and the influences of taphonomic processes and recovery

Table 2.7. Statistical summary of the relationship between NISP and MNI for mammal assemblages from fourteen archaeological sites in eastern Washington State

Site	Regression equation	r	p	$\sum$ NISP	N of taxa
45DO273	MNI = $-0.114(\text{NISP})^{0.58}$	0.816	0.0136	84	8
45OK2A	MNI = $0.01(\text{NISP})^{0.36}$	0.849	0.0019	366	10
45DO282	MNI = $-0.178(\text{NISP})^{0.628}$	0.875	0.0004	426	11
45DO211	MNI = $-0.12(\text{NISP})^{0.64}$	0.847	< 0.0001	474	15
45DO285	MNI = $-0.19(\text{NISP})^{0.51}$	0.721	0.0024	491	15
45DO214	MNI = $-0.07(\text{NISP})^{0.44}$	0.765	0.0003	536	17
45DO326	MNI = $0.02(\text{NISP})^{0.3}$	0.56	0.0242	640	16
45DO242	MNI = $-0.093(\text{NISP})^{0.4}$	0.89	< 0.0001	673	13
45OK287	MNI = $-0.04(\text{NISP})^{0.21}$	0.786	0.007	807	10
45OK250	MNI = $-0.019(\text{NISP})^{0.41}$	0.776	0.003	1,077	12
45OK4	MNI = $-0.072(\text{NISP})^{0.48}$	0.881	< 0.0001	1,108	15
45OK2	MNI = $-0.158(\text{NISP})^{0.4}$	0.769	0.0002	2,574	18
45OK11	MNI = $-0.124(\text{NISP})^{0.5}$	0.849	< 0.0001	3,549	24
45OK258	MNI = $-0.094(\text{NISP})^{0.47}$	0.863	< 0.0001	4,433	21

and identification skills. What we don't know, and can't really know most of the time, is whether MNI is a more accurate measure of taxonomic abundances than NISP for any given assemblage of paleozoological remains. Hudson (1990) and Breitburg (1991) knew that MNI provided more accurate measures of the thanatocoenosis because they knew the original taxonomic abundances of the death assemblage. We don't

Table 2.8. Maximum distinction (each pellet considered independently) and minimum distinction (all pellets considered together) MNI values for six genera of mammals in a sample of eighty-four owl pellets. The right femur is the most abundant element for *Microtus*, and the left femur is the most abundant element for *Peromyscus* in the minimum distinction column

Taxon	NISP	Minimum distinction MNI	Maximum distinction MNI
<i>Sylvilagus</i>	5	1	2
<i>Reithrodontomys</i>	19	5	5
<i>Sorex</i>	46	5	7
<i>Thomomys</i>	68	8	12
<i>Microtus</i>	705	104	118
<i>Peromyscus</i>	1,266	188	220
$\Sigma$	2,109	310	364

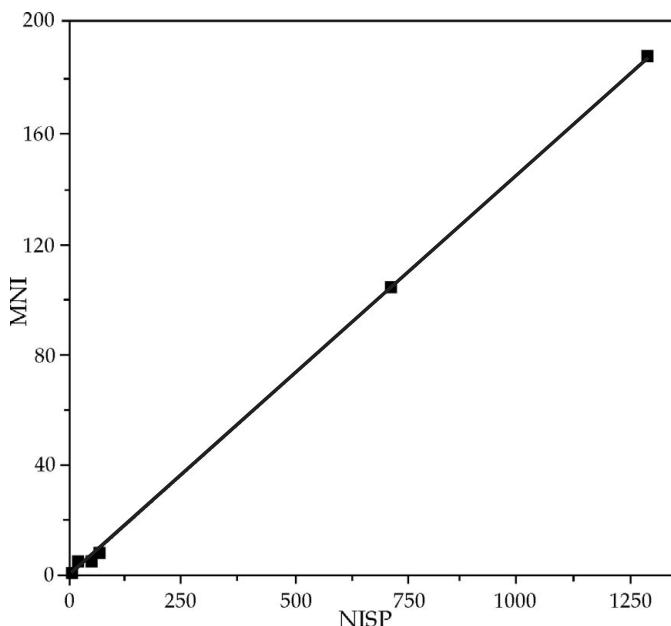


FIGURE 2.8. Relationship between NISP and MNI data pairs for remains of six mammalian genera in eighty-four owl pellets. The regression equation is  $MNI = -0.389(NISP)^{0.149}$ , and the correlation coefficient of the simple best-fit regression line is significant ( $r = 0.999$ ,  $p < 0.0001$ ). Data from Table 2.8.

know what the thanatocoenosis is in paleozoological contexts; it might be our target variable.

That the relationship between NISP and MNI is particularistic and its precise nature is dependent on the samples used is true. But the truth of that claim is not a necessary basis for rejecting NISP in favor of MNI. This is so for several reasons. First, as I noted earlier, NISP often contains virtually the same information regarding taxonomic abundances as does MNI. Second, there are fewer analytical steps in tallying NISP than in deriving MNI, so there are fewer layers (to borrow a metaphor) in the house of cards upon which NISP rests than in the house of cards upon which MNI rests. Do not misinterpret this second point; the simplest method is not being advocated as the best just because it is simpler or easier or contains fewer analytical steps. Rather, because NISP contains fewer steps and, more importantly, fewer assumptions than MNI regarding taphonomy, recovery, identification skills, and the like, perhaps NISP should be preferred. Finally, there is still problem seven, which concerns how to aggregate faunal remains in order to produce MNI values. No such problem exists with NISP, reflecting the fact that NISP is simply additive and that MNI is not additive. It is time, then, to turn to what is likely the most serious problem with MNI.

Table 2.9. Adams's (1949) data for calculating MNI values based on *Odocoileus sp.* remains. MNI-I is equivalent to the MNI minimum distinction (one aggregate, MNI = 118); MNI-II is equivalent to the MNI maximum distinction (five aggregates, MNI = 120)

Recovery provenience	Distal right humerus	Distal left humerus	Distal right femur	Distal left femur	MNI-II
A	38	42	10	11	42
B	5	12	1	1	12
C	30	42	10	10	42
D	11	9	3	5	11
E	9	13	2	7	13
MNI-I	93	118	26	34	

### Aggregation

Although MNI solves the problem of interdependence of specimens inherent to NISP, MNI has its own significant problem. That problem is readily introduced by considering the data that Adams (1949) presented when he determined the MNI of deer (*Odocoileus virginianus*) per recovery provenience unit in the collection he studied (Table 2.9). Adams distinguished what he referred to as "Minimums I" and "Minimums II," the individual column totals and the individual row totals, respectively. I have substituted MNI for "Minimums" in Table 2.9 because MNI is indeed what Adams meant. Notice that were one to ignore recovery provenience, and just tally up the most frequently occurring skeletal part, distal left humeri would be most abundant among the four skeletal parts, so the site-wide MNI – or Adams's "MNI-I" – is 118. But if the analyst were to tally up the most frequently occurring skeletal part per unique recovery provenience, then distal left humeri would be the most abundant skeletal part in four of the five recovery proveniences, but right distal humeri would be the most abundant skeletal part in the fifth recovery provenience. Thus the total MNI for the values summed over the five recovery proveniences – Adams's "MNI-II" – is 120. Grayson (1984) later presented an example from a single site in which differences between MNImni and MNImax values varied across nearly two dozen taxa from 0 to 250 percent (the latter, MNImin = 15 and MNImax = 38).

It makes little difference whether Adams's five distinct recovery proveniences are horizontally distinct (like units in an excavation grid), vertically distinct (as with strata), or both (as with grid units per stratum). His data illustrate the most significant problem that attends MNI. This problem was revealed by Adams (1949:24) when he commented that one must assume "that parts of one individual are not

represented from more than one [recovery provenience].” He said this with specific reference to his “MNI-II” values. But he only revealed the problem; he did not explore its implications. This problem and its implications were later documented at length by Grayson (1973, 1979, 1984). This problem is, in short, known as the aggregation problem, where an aggregate is an assemblage or collection of faunal remains the boundaries of which are chosen by the analyst, whether those boundaries correspond to stratigraphic boundaries or arbitrarily and artificially bounded excavation/collection units.

Grayson (1973) termed what Adams called “Minimums I” values the *minimum distinction method*, and termed what Adams called “Minimums II” values the *maximum distinction method*. The former involves determination of MNI for the complete assemblage considered as one aggregate; the latter involves determination of MNI independently for each assemblage, each from a distinct recovery provenience specified by the analyst. The minimum distinction method is so-called because it produces the lowest or smallest MNI values for a complete collection. The maximum distinction method is so-called because it produces the greatest or largest MNI values for a collection (MNI values for all assemblages from unique recovery proveniences are summed); it produces more than the minimum distinction method because it considers a large number of (small) aggregates (or [sub]assemblages of remains). The minimum distinction method considers only one large aggregate – all remains treated as a single collection. Adams did not care for either the minimum distinction method or the maximum distinction method because, despite the differences in their results, both produced *minimum* numbers of individuals. Furthermore, the maximum distinction method – determining MNI based on individual recovery proveniences – required that one assume specimens in one provenience unit were independent of all specimens in other provenience units, and Adams did not want to make that assumption. It is fitting that we hereafter refer to this potential problem of interaggregate interdependence of skeletal specimens as *Adams’s dilemma*.

That the aggregation problem is widespread is easy to show. Recall the collection of remains of six genera of prey in eighty-four owl pellets; the NISP–MNI data pairs for this collection are plotted in Figure 2.8. That figure is based on the minimum distinction method because all remains were lumped together to form one aggregate. This means that only one skeletal element per taxon contributes to the MNI, regardless of how many pellets contain remains of a taxon. What happens to the MNI values for the taxa represented in the sample of eighty-four pellets when one shifts from the minimum to the maximum distinction method is shown in Table 2.8. The NISP values stay the same regardless of how MNI is determined – whether the maximum or minimum distinction method is used. The MNI is greater in five of six taxa

Table 2.10. *Differences in site total MNI between the MNI minimum distinction results and the MNI maximum distinction results*

Site	N of assemblages	NISP	Richness			N taxa increase	Mean increase per genus
			(N of genera)	MNImin	MNImax		
45OK2A	4	366	10	30	39	4 of 10	0.9
45DO211	4	474	15	108	117	4 of 15	0.6
45DO285	4	491	15	66	102	12 of 15	2.4
45DO214	4	536	17	67	108	11 of 17	2.4
45DO326	4	640	16	53	81	14 of 16	1.75
45DO242	4	673	13	38	52	7 of 13	1.1
45OK250	3	1,077	12	62	79	8 of 12	1.4
45OK4	3	1,108	15	65	82	7 of 15	1.1
45OK2	4	2,574	18	66	105	13 of 18	2.2
45OK11	2	3,549	24	202	231	14 of 24	0.6
45OK258	2	4,433	21	117	139	10 of 21	1.0

when the maximum distinction method is used relative to the minimum distinction method (Table 2.8). The ratio of *Peromyscus* to *Microtus* – the subject of published interpretations of this collection (Lyman et al. 2001, 2003) – shifts from 1.81:1 for the minimum distinction MNI, to 1.86:1 for the maximum distinction MNI, to 1.80 for NISP. In this case, the differences are small, and statistically insignificant; the chi-square value is 0.41 if *Sylvilagus* is omitted so that the assemblage's data pairs meet the requirements of the test ( $p > 0.5$ ). Even given the small differences in ratios of *Peromyscus* to *Microtus*, the critical question is: Which ratio is correct? There is no clear or obvious answer.

The aggregation problem is pernicious. Of the fourteen archaeological assemblages summarized in Table 2.7, eleven have multiple components or temporally distinct (sub)assemblages; the other three consist of only one assemblage. For purposes of generating the regression equations in Table 2.7, I used the minimum distinction MNI values for all fourteen sites. What happens to the MNI values for the eleven sites with multiple (sub)assemblages if the maximum distinction method is used and MNI is derived for each taxon in each (sub)assemblage independently? First, the total MNI for each of the eleven sites increases when one shifts from MNImin(imum distinction) to MNImax(imum distinction) (Table 2.10). Why? Because more kinds of most common elements are specified in the latter than in the former.

Second, the total MNI for each site increases between nine and forty-one when one shifts from MNImin to MNImax; the average increase is 23.7 individuals per

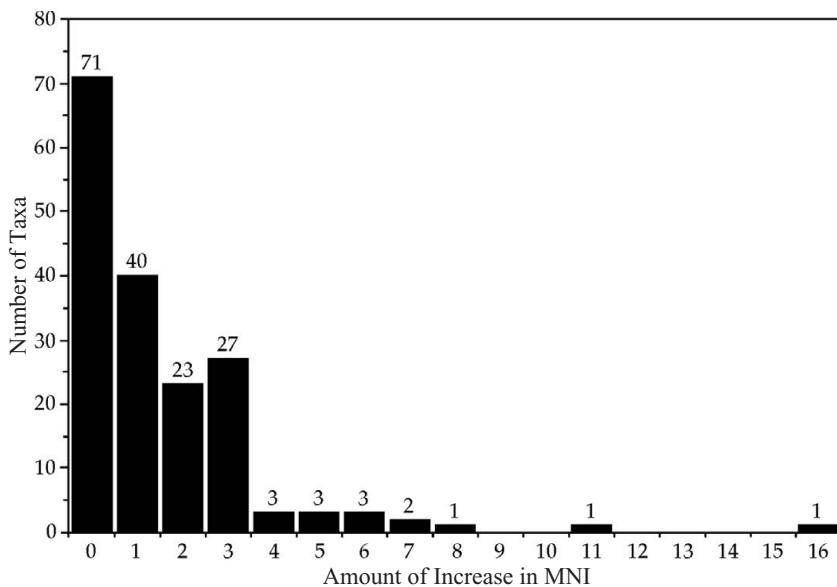


FIGURE 2.9. Amount by which a taxon's MNI increases if the minimum distinction MNI is changed to the maximum distinction MNI in eleven assemblages (see Table 2.10 for other data on these assemblages).

site (Table 2.10). This is not a lot in terms of absolute abundance, but think of it this way: In site 45OK2A, MNImin is thirty and MNImax is thirty-nine; that is a 30 percent increase. In the sample of eleven sites, the site total MNImax increases over MNImin from 8 percent (45DO211) to as much as 61 percent (45DO214); the average increase is a bit more than 35 percent. The third thing to note regarding the shift from MNImin to MNImax is that four to fourteen taxa per site increase in abundance. Not all taxa increase, and any given taxon does not increase consistently in all sites. Ratios of taxonomic abundances shift around rather unpredictably as a result. Note, for example, that the amount by which any taxon's MNI increases is one to sixteen (Figure 2.9). Consider, for example, how the ratio of deer (*Odocoileus* spp.) to gopher (*Thomomys* sp.) changes across all eleven sites when one uses MNImin compared to when one uses MNImax (Figure 2.10). If the MNI of both taxa changed consistently (say, all increase by 10 percent) when shifting from the minimum to the maximum distinction method, the ratios would not change and all points would fall on the diagonal in Figure 2.10. Instead, the eleven collections fall various distances from that line, meaning that the ratios change more in some sites than in others; the changes in most abundant skeletal parts are not patterned. There is marked variation in which skeletal part defines the MNI for either or both deer and gophers across the (sub)assemblages.

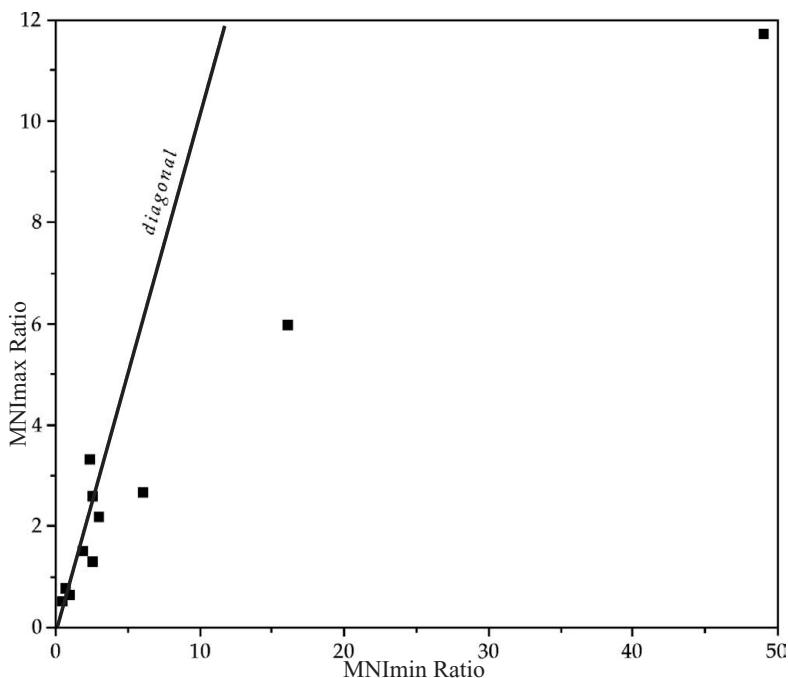


FIGURE 2.10. Change in the ratio of deer (*Odocoileus* spp.) to gopher (*Thomomys* sp.) abundances in eleven assemblages when MNImax is used instead of MNImin. If the ratios did not change, all points should fall on the diagonal line rather than above and to the left, or below and to the right of that line.

The most abundant skeletal part for each of the thirteen mammalian genera at Cathlapotle (Table 1.3) that have more than 1 MNI in each of the two temporally distinct (sub)assemblages are listed in Table 2.11. Only three taxa (of a possible thirteen) have more than one most abundant part (e.g., *Castor* in the postcontact assemblage), indicating rather skeletally uneven representation of individual carcasses. More importantly, it is virtually impossible to predict which part of a genus will be most abundant in one (sub)assemblage based on which part of that genus is most abundant in the other (sub)assemblage, particularly when left and right-side designations are considered (Table 2.11). If the side designation is not considered, then only four genera (*Aplodontia*, *Castor*, *Microtus*, *Ondatra*) out of thirteen are represented by the same skeletal part in both (sub)assemblages. That two of these four particular genera are the ones represented by the same skeletal part regardless of side is, in this case, easy to explain. Only skulls and mandibles of *Microtus* were identified among the mammal remains at Cathlapotle; postcranial remains of this genus were not identified and this markedly increases the probability that the same skeletal part (regardless of side) will be identified in both components. The

Table 2.11. *The most abundant skeletal part representing thirteen mammalian genera in two (sub)assemblages at Cathlapotle. When more than one skeletal part represented the same MNI, all skeletal parts are listed. R, right; L, left*

Taxon	Precontact assemblage	Postcontact assemblage
<i>Lepus</i>	R proximal tibia	R mandible
<i>Aplodontia</i>	L mandible	R mandible
<i>Castor</i>	L femur	R mandible, R femur
<i>Microtus</i>	L mandible	R mandible
<i>Ondatra</i>	L mandible	R mandible
<i>Canis</i>	R P4	L dP4
<i>Ursus</i>	L m2	R ulna
<i>Procyon</i>	R proximal radius	L m1
<i>Mustela</i>	R mandible	R distal humerus
<i>Lutra</i>	R proximal radius	R mandible, R distal humerus
<i>Phoca</i>	L distal humerus	R temporal
<i>Cervus</i>	L astragalus	L naviculo cuboid
<i>Odocoileus</i>	R calcaneum	R m3, R astragalus

mandible of *Aplodontia* is the most frequent skeletal part in both (sub)assemblages because it was selectively retained by site occupants as a wood-working tool – a chisel or engraver (Lyman and Zehr 2003). It is unclear why the same skeletal part provides the MNImax of *Castor* and *Ondatra* in both subsassemblages, but those parts are particularly robust and thus relatively immune to taphonomic attritional processes.

The most frequent skeletal parts in Table 2.11 are from all parts of the skeleton – the head (upper and lower teeth, mandibles), the forelimb, and the hindlimb. It is likely, given what we know about taphonomy at this time, that this is the pattern that will emerge in most cases. Because taphonomic processes influencing the survival and distribution of faunal remains are *not* perfectly correlated with remains that are (or those that are not) taxonomically identifiable (Lyman 1994c), it is unlikely that we will find cases in which the MNImin and MNImax values for a given collection will be perfectly correlated at a ratio scale. They *might* be correlated at an ordinal scale, but it is quite likely even then that the correlation coefficient will be less than 1.0. Shifts in taxonomic abundances will likely not be uniform across all taxa when one shifts from MNImin to MNImax.

Different aggregates of faunal materials making up a total collection will not only produce different MNI values, but they will do so differentially across taxa. Let's say we have one taxon represented in a collection from a site, and that this taxon is represented by twenty-five left and thirty right distal humeri, the most common

*Table 2.12. Fictional data showing how the distribution of most abundant skeletal elements of one taxon can influence MNI across different aggregates. If stratigraphic boundaries are ignored, a minimum of thirty individuals is represented by thirty right humeri. Using stratigraphic boundaries to define aggregates, the total MNI is forty-seven because the most abundant element is left humeri in stratum 1, but the most abundant element in strata 2 and 3 is left humeri*

	Left humeri	Right humeri	MNI per stratum
Stratum 1	22	5	22
Stratum 2	3	17	17
Stratum 3	0	8	8
$\sum \text{MNI}$	25	30	47

skeletal part. Obviously, we have a MNImin of thirty (assuming that we find matches in terms of age, sex, and size for all possible pairs of elements; i.e., the twenty-five left specimens all have matching right specimens). But there are also three strata (or horizontally distinct recovery contexts, if you prefer) comprising the site, and the humeri are distributed across those strata as indicated in Table 2.12. When we sum the MNImax values in Table 2.12, we have a site total of  $\sum \text{MNI} = 47$ . Why? Because whereas with MNImin we had only one most common skeletal part in the form of the right distal humeri ( $\sum = 30$ ), we now have in Stratum 1 the left humerus as the most common part whereas in Strata 2 and 3 the right humerus is the most common part. The change from one kind of most common skeletal part to two kinds resulted in an increase of 17 MNI (57 percent).

As a final example, let's say we have two taxa. Taxon 1 is represented by the remains of 7 individuals (= MNImin); those remains consist of 7 R humeri, 6 L humeri, 6 R femora, and 5 L femora ( $\sum \text{NISP} = 24$ ). Taxon 2 is represented by the remains of 14 individuals (= MNImin); those remains consist of 14 R humeri, 7 L humeri, 6 R femora, and 10 L femora ( $\sum \text{NISP} = 37$ ). If we define faunal assemblages stratigraphically, and there are three strata in the site, we may find the stratigraphic distribution of skeletal parts indicated in Table 2.13. In that table the MNI for taxon 1 shifts from MNImin = 7 to MNImax = 10, and the MNI for taxon 2 does not shift but rather both MNImin = 14 and MNImax = 14. The change in taxon 1 is the result of changes in the number of most abundant skeletal parts defined for this taxon as the aggregates change. Most disconcerting is the fact that the ratio of taxon 1 to taxon 2 changes from 7:14 (or 1:2) to 12:14 (or 1:1.2) with a simple change in aggregates. Again, these changes result from specification of different most common skeletal parts with each different set of aggregates.

Table 2.13. *Fictional data showing how the distribution of skeletal elements of two taxa across different aggregates can influence MNI. If stratigraphic boundaries are ignored, there are only seven individuals (R humeri) of taxon 1, and fourteen individuals (R humeri) of taxon 2. Aggregates defined by stratigraphic boundaries produce twelve individuals of taxon 1 and fourteen individuals of taxon 2*

	Taxon 1	Taxon 2
Stratum 1	6 R humeri, 2 L humeri, 3 R femora, 3 L femora (MNI <sub>max</sub> = 6)	4 R humeri, 1 L humerus, 4 L femur (MNI <sub>max</sub> = 4)
Stratum 2	1 R humerus, 4 L humerus, 1 R femora, 1 L femur (MNI <sub>max</sub> = 4)	4 R humeri, 1 L humeri, 3 R femora, 1 L femur (MNI <sub>max</sub> = 4)
Stratum 3	1 L femur, 2 R femora (MNI <sub>max</sub> = 2)	6 R humeri, 5 L humeri, 4 R femora, 4 L femora (MNI <sub>max</sub> = 6)

Changes like those documented above are likely to occur more often than not. This renders MNI a very unstable measurement unit. Of course, changes in aggregation may not cause MNI values to fluctuate if the distribution of most abundant skeletal parts is the same for each taxon. Consider, for example, the two-taxon fictional data given in the earlier example, but this time with similar distributions of most abundant elements across the three strata, as shown in Table 2.14. The most abundant element of both taxa (R humerus) has a similar distribution across all three strata and displays its greatest frequency in Stratum 1. The ratio of taxon 1 to taxon 2 is 1:2 in all three strata by the MNI<sub>max</sub> method. The ratio of 1:2 is given by the MNI<sub>min</sub> method as well.

Studying the distribution of most abundant elements per taxon across different aggregates may reveal much about site formation and taphonomic history, as Grayson (1979, 1984) noted years ago. I am, however, unaware of any such studies in the literature. This is surprising given interest in *site structure* (e.g., O'Connell 1987). Perhaps the lack of such studies is an instance of benign neglect. Whatever the case, consideration of how aggregation influences MNI deserves more study than it has received because of the insight it will provide to MNI as a measure of taxonomic abundance and also because of the insights it may grant to site structure. In such studies, an aggregate of any kind might be defined – by a site as a whole, by a stratum, by an archaeological feature (each pit, house floor, hearth, etc.), by an arbitrary excavation unit (say, 2 m × 2 m × 10 cm thick), or some combination thereof.

Table 2.14. *Fictional data showing that identical distributions of most common skeletal elements of two taxa across different aggregates will not influence MNI. Note that the NISP per skeletal element is the same as in Table 2.13. Note also that the four distinct skeletal elements have similar frequency distributions across the three strata, and that the ratio of taxon 1 to taxon 2 is 1:2 in each of the three strata, and that MNI values determined while ignoring stratigraphic boundaries also produce a ratio of 1:2*

	Taxon 1	Taxon 2
Stratum 1	5 R humeri, 4 L humeri, 4 R femora, 3 L femora (MNI <sub>max</sub> = 5)	10 R humeri, 5 L humerus, 4 R femur, 8 L femur (MNI <sub>max</sub> = 10)
Stratum 2	1 R humerus, 1 L humerus, 1 R femora, 1 L femur (MNI <sub>max</sub> = 1)	2 R humeri, 1 L humeri, 1 R femora, 1 L femur (MNI <sub>max</sub> = 2)
Stratum 3	1 R humerus, 1 L humerus, 1 R femur, 1 L femora (MNI <sub>max</sub> = 1)	2 R humeri, 1 L humeri, 1 R femora, 1 L femora (MNI <sub>max</sub> = 2)

A final point to consider involves Uerpmann's (1973:311) observation that the "difference between number of finds [NISP] and 'minimum number of individuals' increases as the size of the sample increases" (Uerpmann 1973:311). The difference between NISP per taxon and MNI per taxon will increase as NISP increases (Figure 2.4). Because larger sample sizes allow greater differences between values, differences between MNI<sub>min</sub> and MNI<sub>max</sub> will be greatest in large samples and smallest in small samples. Thus, large sample sizes, which are desired for statistical reasons (large samples tend to be more representative than small samples of the population from which they are drawn, and they tend to increase the statistical power of a test), tend to be the ones in which MNI fluctuates the most as different aggregates are defined. As Grayson (1979:210) noted, "This is not the usual behavior of a unit of measurement."

Restating problem seven, MNI measures not only taxonomic abundances but aggregation methods as well (Grayson 1984). This can be shown graphically and statistically by considering the relationship between NISP and MNI as modeled in Figure 2.4, but with log-transformed data such that the relationship is linear. The slope of the simple best-fit regression line describing the relationship between NISP data and taxonomically corresponding MNI data should be less steep when more agglomerative (larger aggregates) methods are used to calculate MNI (MNI<sub>min</sub>) than when less agglomerative (smaller aggregates) methods are used to calculate

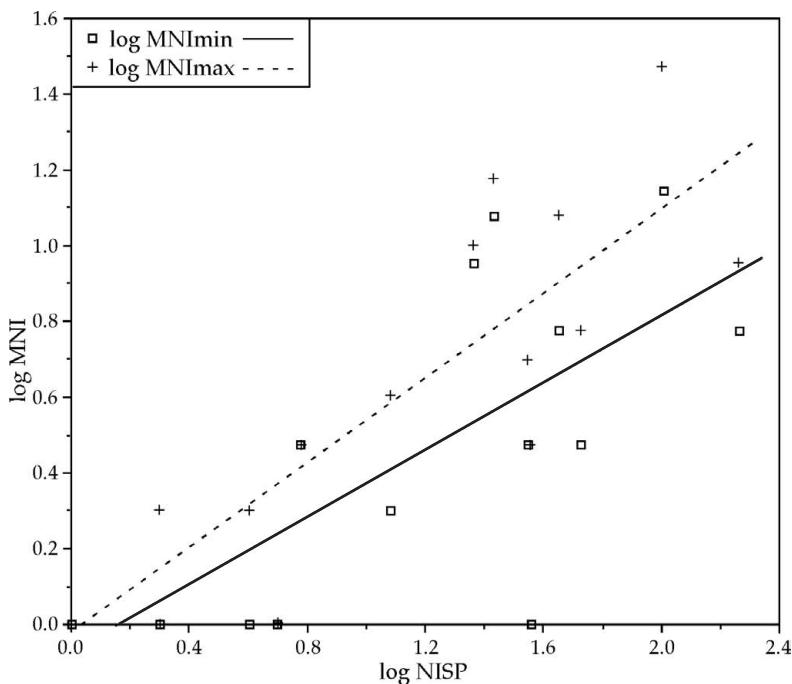


FIGURE 2.11. Relationships between NISP and MNImin, and NISP and MNImax at site 45DO214.

MNI (MNImax). This is so because MNImin involves few most common elements so it is difficult to find and thus add a new most common element. MNImax involves many most common elements so it is easy (relatively speaking) to find and thus add a new most common element. Consider site 45DO214 among the collections from eastern Washington State mentioned previously (Table 2.7). The slope of the simple best-fit regression line describing the relationship between NISP and MNImin is 0.44 (Table 2.7); the slope of the simple best-fit regression line describing the relationship between NISP and MNImax is 0.57. Both sets of data points and both best-fit regression lines are included in Figure 2.11, which shows that the slope for MNImin is less steep than that for MNImax.

The example of 45DO214 indicates that MNI not only measures taxonomic abundances but it also measures aggregation. But we do not want measures of two variables to obscure each other, particularly when one may approximate the target variable and the other has nothing whatsoever to do with the target variable. Is there, perhaps, a logical way to define aggregates such that the measures of taxonomic abundances provided by MNI can be treated as if the aggregates do not significantly influence those abundances?

### Defining Aggregates

Recall that an *assemblage* of faunal remains is the set of remains from a horizontally and vertically bounded space, usually a geological space such as a stratum or a part thereof. Following Grayson (1984), the term *aggregate* is used as a synonym for assemblage, and the term *aggregation* for the process of defining the spatial boundaries of a faunal assemblage. Despite Grayson's (1973) recognition of the aggregation problem more than 30 years ago, few analysts other than Grayson (1979, 1984) have subsequently explored its implications with their own data. Thus it is not unusual to find paleozoologists still calculating MNI values without considering the aggregation problem (e.g., Trapani et al. 2006). The aggregation problem is not even mentioned in one textbook on zooarchaeology (Rackham 1994).

Payne (1972) suggests that aggregates of faunal remains should be defined on the basis of the homogeneity of taxa and their frequencies. Ignore for the moment the question of how similar is similar enough for two assemblages to be considered homogeneous (Payne did not address this question), and consider the following three things. First, this procedure assumes that *natural* faunal aggregates exist and we have but to discover them. But whether natural discoverable faunal aggregates exist or not is unclear. Furthermore, what kinds of faunal aggregates are to be searched for – those representing a depositional event, a human-behavior, a death event, or . . . what? Perhaps the research question being asked would help specify the appropriate aggregate, but other problems attend their definition. The second thing to consider, then, is that Payne's protocol precludes the study of stasis because one aggregates, say, stratigraphically sequent, similar (homogeneous) assemblages. Finally, Payne's procedure comprises a circular process: Change in the fauna would be identified based on how the aggregates were defined – the property of differences or nonhomogeneity interpreted as change – because aggregates are defined on the basis of similarity and homogeneity. Based on these three observations, we might use nonfaunal criteria to define aggregates.

Ringrose (1993:128) argues that “not all levels of aggregation are likely to be sensible, so that the problem [of the influence of aggregation on MNI] is perhaps less than it might seem at first. If it is not possible for specimens from the same individual to be present in two locations [that is, to be tallied in two distinct aggregates], then it is nonsensical to calculate the MNI at a level of aggregation where these two locations are taken together, since specimens will be, implicitly, counted as being possibly from the same individual when in fact they cannot be.” Implementing this protocol of defining aggregates demands a great deal of knowledge regarding the taphonomic history of the materials under study. Some of it might be found by refitting studies

(e.g., Rapson and Todd 1992; Todd and Stanford 1992). However, this is again using faunal data to define aggregates and thus imparts a degree of circularity to those definitions. It also can introduce the problem of matching potentially paired (left and right) skeletal parts (e.g., Todd and Frison 1992).

Some zooarchaeologists indicate that one should define faunal aggregates based on “cultural units rather than arbitrary ones related to excavation logistics” (Reitz and Wing 1999:198). This is all well and good, but does this mean that two pits containing bones and originating in the same stratum (dating to the same time period and apparently representing the same cultural context given stratigraphic contemporaneity) should be considered separately or together? Archaeologist James Ford (1962) argued long ago that archaeological “cultural units” such as cultures, phases, periods, and the like were often defined on the basis of stratigraphically bounded aggregates of artifacts, but that it was unclear why there should be any necessary relationship between sediment deposition boundaries and boundaries between cultures. I agree (Lyman and O’Brien 1999). So what are we to do?

Let us begin by glancing at the solution that paleontologists have used. Fagerstrom (1964:1198), a paleontologist interested in past biological communities, suggested that a *fossil assemblage* representing a community was “any group of fossils from a suitably restricted stratigraphic interval and geographic locality.” What is suitable is not clear, though it is hinted at in other paleontological concepts. In vertebrate paleontology, a *faunule* is an assemblage of associated animal remains from one or several contiguous strata, dominated by members of one biological community (Tedford 1970:677). And, a *local fauna* is a set of remains from one locality or several closely spaced localities which are stratigraphically equivalent or nearly so, thus it is a set of taxa close in (geological) time and (geographic) space (Tedford 1970:678). Identifying prehistoric faunal communities – or faunules – was what Shotwell (1955, 1958) was concerned about, and he emphasized the taphonomic problems with doing so when one used an aggregate of fossils the boundaries of which were set by excavation strategies and stratigraphy.

The preceding brief discussion hints at two things. First, paleontologists often seek, like Chester Stock and Hildegard Howard did, to determine the census of a paleocommunity, or a biocoenose. That is their target variable, and they acknowledge the geological mode of occurrence of the faunal materials that they study, and they use stratigraphic boundaries and extent of exposures to collect a sample of those materials. The second thing hinted at is an extremely critical detail. Reitz and Wing (1999:197) mention it when they state that the aggregates of faunal remains defined “may depend on the research problem.” Any aggregates defined *must* depend on the research problem, as well as whatever taphonomic and site-formational information is available. Thus, on the one hand, human behaviorally significant assemblages of

remains, such as those in cache pits or in trash middens or on house floors are likely to be important to questions about human interactions with fauna. On the other hand, questions regarding paleoecology are likely to be phrased in such a manner as to require temporally and spatially distinct assemblages of remains, perhaps but not necessarily representing one “biological community” but certainly providing insights to the nature of biocoenoses. Temporally distinct assemblages but perhaps not human behaviorally significant ones would be of interest to paleoecologists.

Valensi (2000:358) noted that aggregation based on excavation levels “gave an overestimation of MNI [as a result of specimen] interdependence [across] some levels.” Interdependence was recognized by refitting specimens of both lithic and bone specimens that came from different depositional units. Valensi used archaeostratigraphic units as the basis for defining aggregates, and found refitting specimens that came from different units. Her analysis suggests a protocol for defining aggregates. Refits of lithic specimens would provide nonfaunal criteria for defining faunal aggregates. The paleozoologist could adopt a rule, such as only when refits across aggregates are minimal, whereas refits within aggregates are maximized, have appropriate aggregates for determining MNI been defined. However, not only is the time cost incredibly high if the assemblage is large – do the faunal refits, using the same rule, define the same aggregates as the lithics (or ceramics)?

Research questions about taphonomic histories likely will require an estimate of a taphocoenose, those about hunter or predator selectivity will require not only an estimate of a thanatocoenose but also the biocoenose from which it derived. Explicit statement of the research problem and research questions should help the paleozoologist define aggregates that are pertinent. Of course, any available taphonomic information such as obvious refits should also be consulted to help set geological spatial boundaries around the aggregate(s). This does not mean that one will automatically have aggregates that do not share specimens from the same individual, but perhaps those will be so rare as to not significantly bias any statistical results.

## DISCUSSION

Thus far the problems with NISP as a quantitative unit giving valid measures of taxonomic abundances (even in a taphocoenose, let alone in a thanatocoenose or biocoenose) have been considered and it has been argued that all but one of those problems – that of possible specimen interdependence – can be fairly easily resolved analytically. (Some analysts still fail to realize how easily many of the problems with NISP can be resolved analytically [e.g., O’Connor 2001].) Problems with MNI as a quantitative unit giving valid measures of taxonomic abundances have also been

identified and discussed, and it has been shown that many of those are also readily dealt with analytically. The problem that remains with MNI is aggregation. As implied above, there is no magic algorithm for solving the aggregation problem because each aggregate specified by the analyst may, or may not, have a set of faunal remains all of which are indeed independent of all other faunal remains in all other aggregates. Earlier I referred to the latter as *Adams's dilemma*. It is aptly referred to as a dilemma because if, say, stratigraphically bounded aggregates are chosen as the assemblages to be analyzed, one must assume that the faunal remains in each are independent of all other faunal remains in other aggregates. But, of course, they might not be.

A chosen sampling design may indicate where to excavate and which screen-mesh size to use, but the faunal specimens recovered are a result of the taphonomic history of the assemblage – which bones and teeth were accumulated, deposited, and still exist, and where they are located, both horizontally and vertically. The existing remains of a single individual may be in one or more horizontal loci, in one or more vertical loci (or strata), or both. (No two specimens can, of course, occupy exactly the same horizontal and vertical position. By *same location*, I mean a spatially limited, horizontally and vertically bounded unit.) Even attempting to match and pair all skeletal specimens from all excavated recovery proveniences, we will likely never know what the *correct* aggregates of faunal remains should be. By *correct* is meant those that are not only relevant to our research questions, but also ones defined such that specimens from a single carcass are *not* distributed across two or more aggregates. Given that we cannot know this, we either assume Adams's dilemma does not exist, or, we do something other than determine MNI values.

There is, in fact, a relatively simple solution to Adams's dilemma. The solution rests on the fact that quite often, virtually the same information regarding taxonomic abundances in an assemblage is found in NISP as is found in MNI. This statistical relationship has been known for some time (Casteel 1977, n.d.; Grayson 1978a, 1979). In short, MNI is redundant with NISP, where "redundant" means that the two quantitative units produce the same information. The "same information" can mean identical, or simply statistically indistinguishable. To show that MNI and NISP provide the same information in both of these senses, consider the owl pellet data mentioned before. Recall that the sample comprises eighty-four pellets, that the relationship between NISP and MNImin is linear (Figure 2.8), and that the relationship is strong ( $r = 0.989$ ,  $p < 0.0002$ ). For this sample, 97.8 percent of the variation in MNI values is explained by variation in NISP values. Clearly, MNI is redundant with NISP. And, the same applies to the fourteen samples of mammal remains from eastern Washington State (Table 2.7). For these assemblages, the relationship between NISP and MNImin is typically strong ( $r > 0.75$  for 13 of the 14) and significant ( $p < 0.01$  for all). For thirteen of these fourteen assemblages, NISP accounts for

more than 51 percent of the variation in MNI. MNI provides information about taxonomic abundances that is redundant with that provided by NISP (Figure 2.4). But so what? Constructing an answer to this question requires a consideration of the scale of measurement represented by NISP and by MNI.

### WHICH SCALE OF MEASUREMENT?

Some years ago, Grayson (1984:94–96) noted several critical things. First, he noted that converting from one ratio scale to another ratio scale based on different measurement units will not alter the value of a ratio of measurements. This is so because both ratio scales have natural zero points and their respective units of measurement stay constant in each. Thus, the ratio of the weight of two items will not alter if first measured in pounds and then in kilograms. If the two items are 50 pounds and 75 pounds, the ratio of their weights is 1:1.5; the two items weigh 22.68 kilograms and 34.02 kilograms, respectively, for a ratio of 1:1.5. As noted earlier in this chapter, aggregation has the unsavory characteristic of altering MNI tallies, thereby causing ratios of taxa to change as the manner in which faunal remains are aggregated changes.

The second thing Grayson (1984) noted was that MNI values are not ratio scale values precisely because they are *minimum* numbers (Table 2.4). The *actual* number of animals represented (by the identified assemblage) could be as great as the NISP, although it likely will fall somewhere between the MNI and the NISP given the probability ( $> 0.0$ ) of some interdependence of specimens. Thus it cannot be argued that a taxon represented by an MNI of ten is half as abundant as a taxon represented by an MNI of twenty, nor can it be argued that if two taxa each have MNI values of fifteen they are equally abundant. Figure 2.12 plots ratios of the abundances of each pair of taxa based on NISP, MNImin, and MNImax measures of the four least common taxa in the collection of eighty-four owl pellets (Table 2.9). The ratios vary by greater or lesser amounts across the three quantitative measures. In particular, note the variation in ratios between the MNImin and MNImax values. There is no way to determine which set of ratios most closely measures the actual abundances of taxa. Clearly, it is ill-advised to treat MNI values as ratio scale because there are many reasons why they likely are not.

Unfortunately, it is unlikely that NISP values are ratio scale. As Grayson (1984) noted, if MNI provides *minimum* tallies, NISP provides maximum tallies. Given that we do not know the nature (extent) of interdependence of the specimens comprising the NISP for any given taxon in any given collection, and given that intertaxonomic variation in such interdependence will differentially influence how closely NISP tracks a taxon's actual abundance, it is unlikely that ratios of taxa based on NISP values

Table 2.15. Ratios of abundances of pairs of taxa in eighty-four owl pellets. Original data from Table 2.8. Taxon 1, *Sylvilagus*; taxon 2, *Reithrodontomys*; taxon 3, *Sorex*; taxon 4, *Thomomys*

Taxon pair	NISP	MNImin	MNImax
1–2	0.26	0.20	0.40
1–3	0.11	0.20	0.29
1–4	0.07	0.12	0.17
2–3	0.41	1.00	0.71
2–4	0.28	0.62	0.42
3–4	0.68	0.62	0.58

are in fact ratio scale. There is no way to know which set of ratios of abundances of taxa in the owl pellet fauna (Table 2.15), if any, most accurately reflects the actual ratio scale abundances of the taxa. As Grayson (1984:96) noted, because “we know nothing of the nature of the frequency distribution [of taxonomic abundances] that begins with MNI and ends at NISP for a set of taxa,” knowledge of ratio scale abundances of taxa is precluded.

If MNI and NISP do not provide ratio scale taxonomic abundance data, do they perhaps provide ordinal scale abundance data? Again, Grayson (1984:96–99) provided

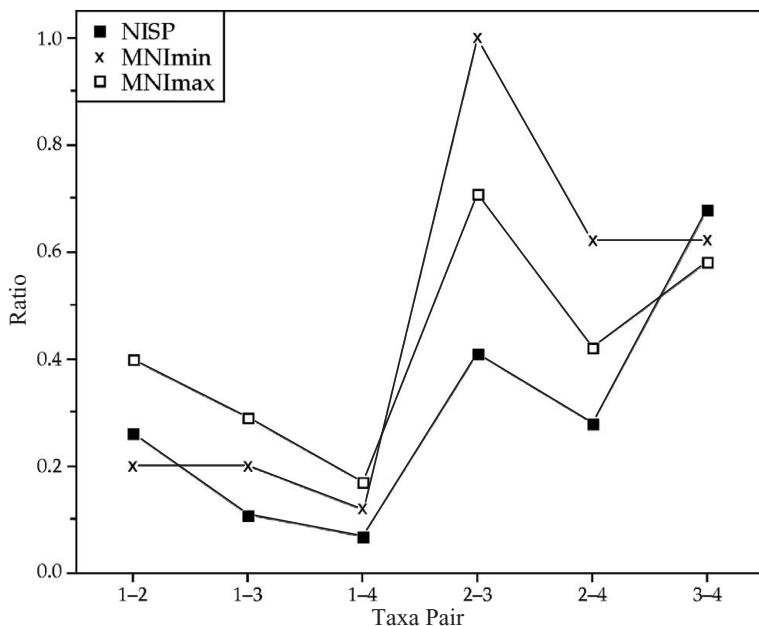


FIGURE 2.12. Ratios of abundances of four least common taxa in a collection of eighty-four owl pellets based on NISP, MNImax, and MNImin. Taxon 1, *Sylvilagus*; 2, *Reithrodontomys*; 3, *Sorex*; 4, *Thomomys*. Data from Table 2.8.

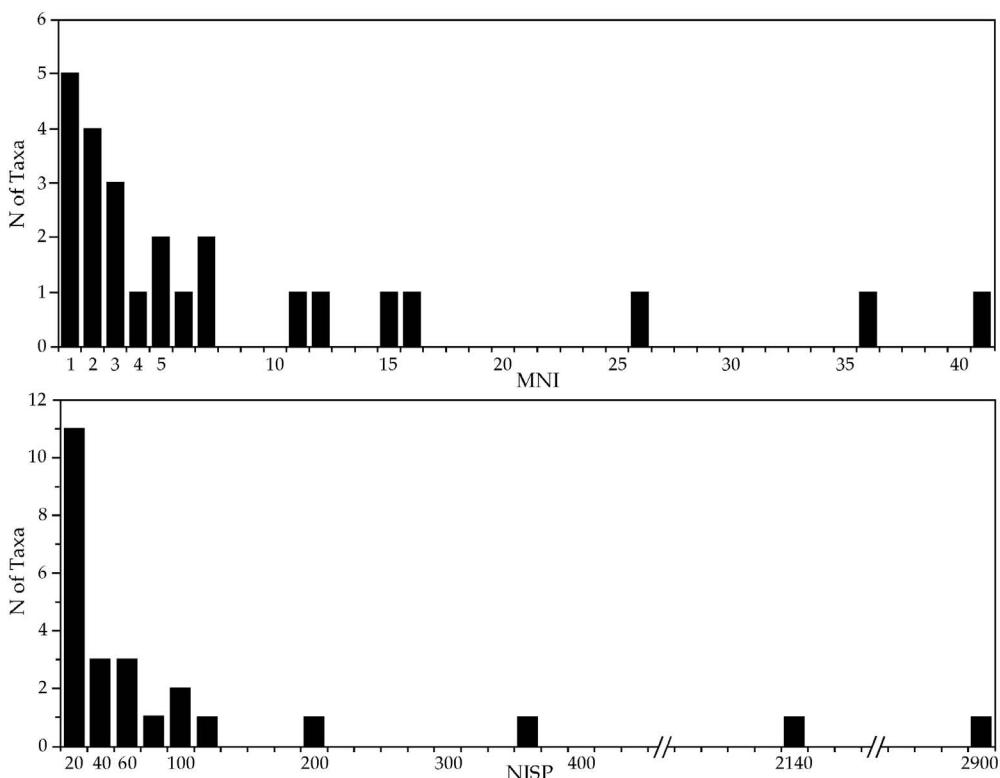


FIGURE 2.13. Frequency distributions of NISP and MNI taxonomic abundances in the Cathlapotle fauna. Data from Table 1.3.

a clear answer. The rank order abundances of taxa are often quite similar across NISP and MNI; if the two sets of values are significantly correlated on an ordinal scale, then the included taxonomic abundances are ordinal scale. Why NISP and MNI should often be correlated comprises the critical insight as to why we can conclude they are ordinal scale. In most multitaxa faunas, a few taxa are represented by many individuals and specimens, and many taxa are represented by few individuals and specimens. As taxonomic abundances increase (whether NISP or MNI), the magnitude of the differences between abundances of adjacent taxa increases. Such frequency distributions increase the probability that taxonomic abundances are ordinal scale because there is less chance that variation in aggregation (MNI) or specimen interdependence (NISP) will alter rank order abundances.

Summing the precontact and postcontact assemblages, eighteen taxa in the Cathlapotle fauna (Table 1.3) are represented by seven or fewer individuals whereas only seven taxa are represented by more than ten individuals (Figure 2.13). Similarly, 20 taxa are represented by 100 or fewer specimens whereas only 5 taxa are represented by

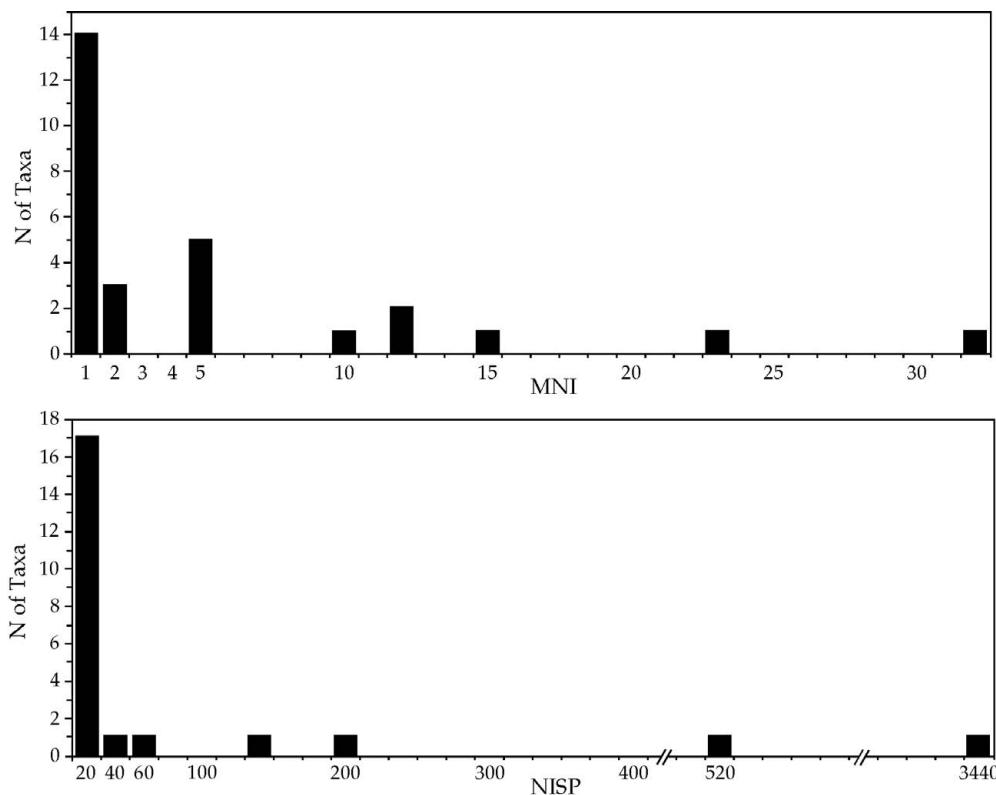


FIGURE 2.14. Frequency distributions of NISP and MNI taxonomic abundances in the 45OK258 fauna in eastern Washington State.

more than 100 specimens. One of the mammalian faunas from eastern Washington State that has been used in other analyses – site 45OK258 – has similar frequency distributions (Figure 2.14). Grayson (1979, 1984) presented numerous other faunas with right skewed distributions of taxonomic abundances. Lest one think this sort of frequency distribution is a function of the faunas we examined, two more examples may convince the skeptic. A right skewed frequency distribution is found in the two summed late prehistoric assemblages of mammal remains from the western Canadian Arctic described by Morrison (1979) (Figure 2.15). And such a frequency distribution is also found among historic era mammalian faunas described by Landon (1996) (Figure 2.16).

It matters little why many faunas display a right skewed frequency distribution of taxonomic abundances [Box 2.2]. The important point is, as Grayson (1984:99) put it, the “degree of rank order stability will be closely related to the degree of

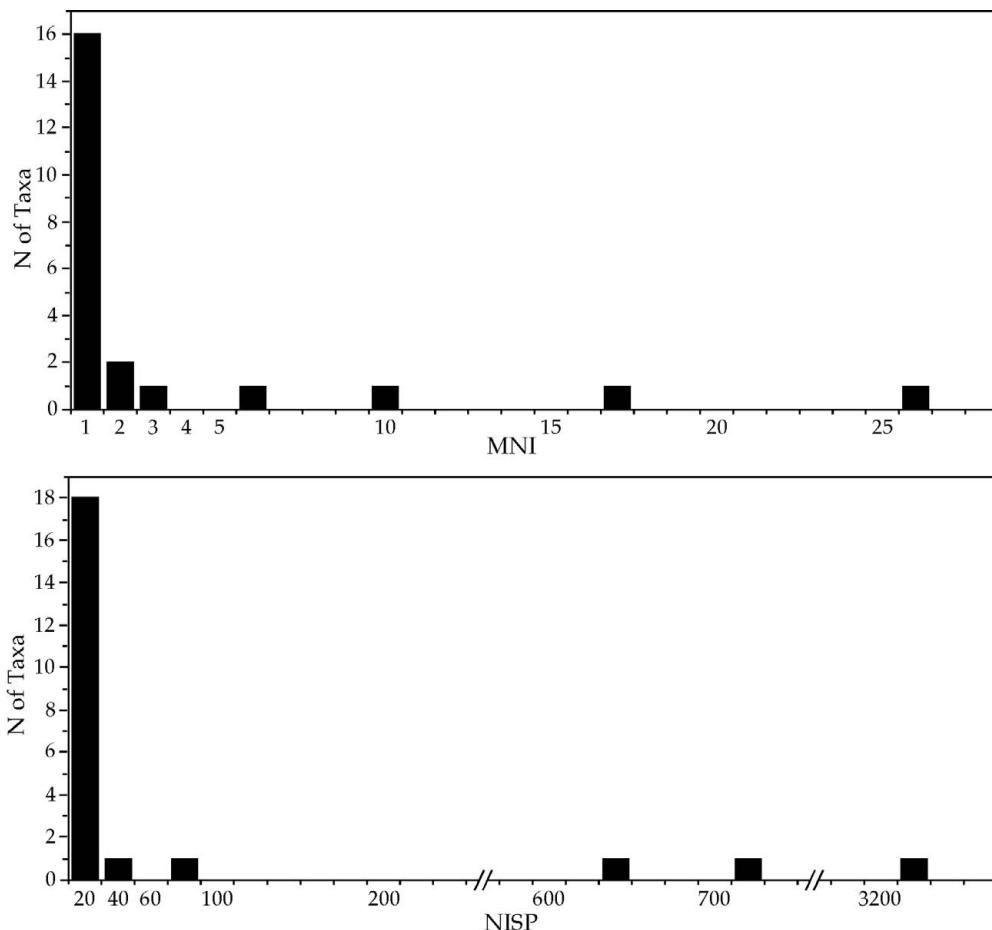


FIGURE 2.15. Frequency distributions of NISP and MNI taxonomic abundances in two lumped late-prehistoric mammal assemblages from the western Canadian Arctic. Data from Morrison (1997).

separation of the taxa involved in terms of their MNI- or NISP-based sample sizes.” The greater the separation between the abundance of taxon A and the abundance of taxon B, the less likely changes in aggregation – if abundances are MNI values – or specimen interdependence – if abundances are NISP values – will alter the rank order abundances of those taxa. The rank order abundances of rarely represented taxa, because their abundances are not widely separated, likely will shift with changes in aggregation and specimen interdependence. As Grayson (1984:98) suggests, “it is questionable whether [rarely represented taxa] should be treated in anything other than a nominal, presence/absence sense.”

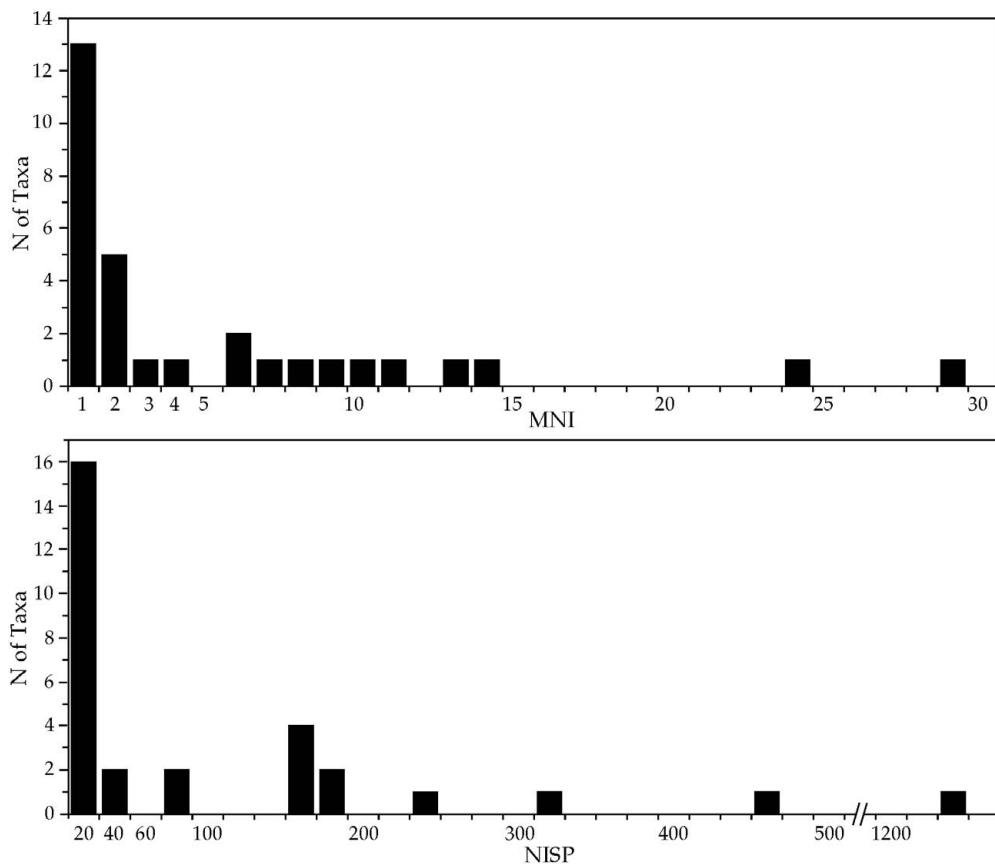


FIGURE 2.16. Frequency distributions of NISP and MNI taxonomic abundances in four lumped historic era mammalian faunas. Data from Landon (1996).

### Box 2.2

This sort of frequency distribution is known as *right skewed* – the tail is to the right. Such a frequency distribution may result from the accumulation agent focusing on one or a few taxa – the frequently represented ones – and the rarely represented taxa are background or idiosyncratic accumulations. Alternatively, recovery may create the frequency distribution. This would be suggested by rare remains of small animals and frequent remains of large animals (see Chapter 4). Finally, perhaps the frequency distribution represents what is on the landscape if accumulation, preservation, and recovery were all random with respect to the biocoenose.

One can determine if taxonomic abundances are ordinal scale – that their rank order of abundance does not alter with counting method – by calculating the rank order correlation between taxonomic abundances produced by the most agglomerative approach to quantification (MNI<sub>min</sub>) with the most divisive approach to quantification (NISP). In most cases, there will be a limited number of ways to aggregate faunal remains (e.g., by site, by stratum, by excavation unit, from most to least agglomerative). If the rank orders of abundances indicated by the most and by the least agglomerative methods are significantly correlated, then the effects of aggregation, of specimen interdependence, or both are such that they do not influence the ordinal scale abundances of taxa (Grayson 1984:106). Analysis and statistical manipulation of the rank ordered abundances is appropriate. Grayson (1979:216, 1984:98) cautioned, however, that aggregation will tend to most strongly influence the rank ordered MNI abundances of rarely represented taxa precisely because they are rare and thus there are minimal to no gaps between their absolute abundances (MNI of one vs. two, say). Changes in aggregation will cause shifts in taxonomic absolute abundances and thus changes in rank ordered abundances, especially increasing or decreasing the number of tied ranks (because there are few ranks of rare taxa, say one–three or four, yet typically a half dozen or more taxa in those ranks). A similar argument applies to NISP and interdependence. Rarely represented taxa are more likely to shift rank orders of abundance than abundant taxa if interdependence could be validly determined because the abundances of rare taxa will not be separated by large gaps in abundance, whereas abundant taxa are likely to be separated by large differences in abundance.

If the rank order abundances fluctuate across different tallying methods such that NISP and MNI<sub>min</sub> are not correlated, then one must conclude that the data are at best nominal scale. Taxa must be treated not as quantitative variables, but as qualitative variables or attributes of a collection; taxa must be treated simply as present in or absent from the collection under study. The analyst could initiate a detailed taphonomic study in an effort to determine if such things as intertaxonomic variation in fragmentation are influencing analytical results. Alternatively, a qualitative interpretation of the taxa present would be reasonable, such as saying that the prehistoric occupants of the archaeological site that produced the faunal remains ate various taxa, but not delving into whether more of taxon A or more of taxon B was eaten. Given that the greatest changes in rank ordered abundances will be in rarely represented taxa, it is the rare taxa that likely should be treated as nominal scale. How rare is “rare” in any given assemblage is an empirical issue. The paleozoologist could initiate an exploration of how rare is rare by generating a graph like those in Figures 2.13–2.16, paying attention to which taxa fall to the left and have no gaps between their abundances.

If NISP and MNI both might produce ordinal scale data on taxonomic abundances, which should be used? I have argued that the definition of aggregates is a serious problem given minimal logical consideration by most paleozoologists, yet the aggregates defined will typically have an influence of greater or lesser magnitude on MNI values. I have also pointed out that MNI is a derived measure and as such MNI values will be influenced by the attributes the analyst chooses to assess specimen interdependence, such as size, ontogenetic age, and sex. One who prefers MNI might find the problems of aggregation and derivation to be minor ones relative to those associated with NISP, but NISP is additive regardless of aggregation; MNI is not. NISP is influenced by specimen interdependence but MNI is not (at least, not as much, remembering Adams's dilemma). Acknowledging the combination of difficulties attending each quantitative unit, which should be used? Given that NISP is redundant with MNI, the answer seems obvious. Use NISP to determine taxonomic abundances.

## RESOLUTION

MNI is at best an ordinal-scale measurement unit. NISP is likely to provide an ordinal-scale measure of taxonomic abundances at best. MNI underestimates the true abundance of some taxa, overestimates others, and accurately estimates the abundances of still others. NISP is likely to do the same, although the taxa the abundances of which are overestimated may not be the same as those that are overestimated by MNI, and so on. And, if MNI provides minimum values and as a result cannot provide mathematically valid ratios, then because NISP provides maximum values it cannot provide mathematically valid ratios either. Finally, recall the tight statistical relationships found between the NISP and the MNI (usually MNImin) evident in many assemblages (Tables 2.6 and 2.7; plus assemblages described by Bobrowsky [1982], Casteel [1977, n.d.], Hesse [1982], Grayson [1979, 1981b], and Klein and Cruz-Uribe [1984]). That relationship suggests that if MNI is at best an ordinal-scale measure of taxonomic abundances, so, too, is NISP. The argument can be made in reverse order; if NISP is ordinal scale and is correlated with MNI, then MNI is also ordinal scale.

Do not be confused by the argument that NISP is at best an ordinal-scale measure of taxonomic abundances. Figures 2.5–2.8 and 2.11, and Tables 2.6 and 2.7 all treat NISP data as if they are ratio scale, but note that no ratio scale interpretations of those data have been offered. Were the owl pellet data in Table 2.9 and Figure 2.8 to be interpreted in ratio scale terms, one might say that *Reithrodontomys* was nearly

four times as abundant as *Sylvilagus* based on NISP, or five times as abundant based on MNImin. NISP and MNI data for the six genera in the eighty-four owl pellets are presented in ratio scale terms, but are interpreted in ordinal-scale terms (Lyman and Lyman 2003). This is not an unusual analytical protocol. It is, for example, typical of how palynologists operate; they present ratio scale data on abundances of plant taxa in a pollen diagram and interpret those data in ordinal scale terms (Moore et al. 1991). The reasons for this protocol are similar to those in paleozoology. There is, for example, intertaxonomic variation in the rate of pollen production, intertaxonomic variation in the accumulation rate (and transport distance) of pollen, and the like. Palynologists realize that counting pollen grains will produce ratio scale counts but that those counts are best interpreted in ordinal scale terms. With respect to paleozoological data, MNI and NISP are both typically at best ordinal scale measures of taxonomic abundance, and they are correlated, often rather strongly. The information on taxonomic abundances provided by MNI is also often provided by NISP.

NISP is a fundamental measurement whereas MNI is a derived measurement. The only analyst-related source of variation in NISP involves identification skills. That source of variation is joined by other analyst-related sources when MNI is calculated. How pairs of left and right elements are sought by an analyst can vary. Does the analyst doing the matching consider only size, only shape, only ontogenetic age? Are the specimens matched visually as when all left femora are laid out on the lab table and compared with all right femora, or are verbal descriptions compared? And there is always the potential for interobserver variation even if precisely the same methods of matching are used. Finally, it is unclear if two analysts will define precisely the same aggregates even if they have the same research question. Despite such issues, paleozoologists continue to try to design valid ways to derive MNI values (e.g., Avery 2002; Vasileiadou et al. 2007).

In light of the discussion to this point, one conclusion seems inescapable: Why bother with MNI when NISP is more fundamental, less derived, and the two provide redundant information? True, NISP *seems* to have more problems than MNI, but many of the problems with NISP are easily dealt with analytically or concern interdependence. Fragmentation, for example, increases NISP to some degree; rather than tally one skeletal element in an assemblage with broken skeletal elements, two or more pieces (specimens) of the same element are tallied. The same applies to intertaxonomic variation in differential transport of skeletal parts and portions, intertaxonomic variation in numbers of identifiable elements, and the like. Such criticisms of NISP not only reduce to concerns about specimen interdependence,

they seem to originate specifically from the perspective that an *individual organism* is the proper counting unit, regardless of anything else. Recall that MNI seems to be commonsensical to calculate and that it has a basis in empirical reality because of the individuality of every organism. But empirical verifiability of individuals is a weak warrant to use individuals as the quantitative unit in paleozoology, especially when it is recognized that bones and teeth are also empirically verifiable biological units. And, just because much of biology focuses on individual organisms or multiples thereof, should paleozoology adopt that focal unit? The answer to that question depends on the research problem under investigation and the attendant target variable(s).

If one adopts the argument that NISP should be the preferred unit with which to measure taxonomic abundances, then there remains the potential problem of specimen interdependence that plagues NISP. As Grayson (1979, 1984) has argued, that problem is rather easily dealt with also. He noted that the “effect of interdependence upon specimen counts is much the same as that of aggregation on minimum numbers: both have the potential of differentially altering measured taxonomic abundances” (Grayson 1979:222). Grayson argued that aggregation will not differentially alter MNI if, and this is a critical *if*, most abundant specimens are identically distributed across aggregation units (Table 2.14 and associated discussion). Similarly, Grayson (1979:223) noted that the interdependence of specimens should not significantly influence NISP as a measure of taxonomic abundances if, and again this is a critical *if*, “all specimens are independent of one another,” or “interdependence is randomly distributed across taxa.” The former is unlikely, and there is no well established method for determining whether or not specimens are independent of one another, or even if they are truly interdependent. How do we determine if interdependence is randomly distributed across taxa?

MNI is a function of NISP; if we know the NISPi values for an assemblage we can typically rather closely predict what the MNIIi values for that assemblage will be. And note that although MNI measures both taxonomic abundances and aggregation method, it does provide what are likely to be independent values, especially if we determine MNI<sub>min</sub> in order to avoid Adams’s dilemma that skeletal parts in different aggregates may not be independent. Finally, note that NISP values are likely to be interdependent to some degree. Putting these observations together, it seems logical to conclude that if MNI<sub>min</sub> and NISP are correlated, then we can assume that interdependence of identified specimens is randomly distributed across taxa because MNI<sub>min</sub> is not influenced by interdependence. In conjunction with the fact that MNI is redundant with NISP, there is little reason to use MNI as a measure of taxonomic

abundances. NISP will work nicely as a unit with which to measure taxonomic abundances at an ordinal scale.

## CONCLUSION

NISP is to be preferred over MNI as the quantitative unit used to measure taxonomic abundances. Throughout much of this chapter, the target variable has been referred to as *taxonomic abundances*, with only occasional reference to whether those abundances pertained to a biocoenose, thanatocoenose, taphocoenose, or identified assemblage. It should be clear, however, that what is measured by either MNI or NISP concerns the set of materials lying on the lab table. The taxonomic abundances are most directly related to the identified assemblage, less directly to the taphocoenose, even less directly to the thanatocoenose, and least directly to the biocoenose from which the remains derived in the first place. It is in part for this reason that at least one alternative method – that of matching paired bones discussed in Chapter 3 – was proposed. A more direct measure of the thanatocoenose was desired, but whether or not such is actually attained is debatable.

Aggregate definition must depend on the research question being asked. That question should explicitly state the target variable(s), and it should be identified as the identified assemblage, the taphocoenose, the thanatocoenose, or the biocoenose. Grayson (1979, 1984) seldom mentioned which of these potential target variables was of interest, though his substantive analyses at the time suggest he sought a measure of taxonomic abundances within the biocoenose. Grayson was particularly worried about the statistical and mathematical properties and relationships of NISP and MNI. This concern is reflected by his focus on the effects of aggregation and of interdependence. But many other analysts also failed to make explicit which one (or more) of the potential target variables was of interest. It is in part for this reason – the lack of an explicitly specified target variable – that many paleozoologists, especially zooarchaeologists, have argued for decades about how to determine taxonomic abundances. There is not nearly as extensive a literature on this topic in paleontology, which is not to say that there are not titles on this topic in the paleontological literature (e.g., Badgley 1986; Gilinsky and Bennington 1994; Holtzman 1979). The reason that there is not as extensive a literature in paleontology as there is in zooarchaeology is because the former generally has one and only one target variable, and it is the same regardless of researcher. That target is the biocoenose. Zooarchaeologists, on the other hand, often have rather different target variables depending on the

questions they are asking. What did people eat versus what was available to exploit, for example.

Explicitly specifying the exact target variable will go a long way toward clarifying an appropriate (valid) quantitative unit. It is exactly such specification that prompted some researchers to develop and use methods of measuring taxonomic abundances other than NISP and MNI. We turn to those alternative units in Chapter 3, and then in Chapter 4 we return to NISP and MNI to explore how they have been and can be used to measure properties of prehistoric faunas.

## Estimating Taxonomic Abundances: Other Methods

In Chapter 2, the two methods of measuring taxonomic abundances – NISP and MNI – most commonly used in paleozoology were discussed. In this chapter other methods that have been used to quantify taxonomic abundances or what is sometimes loosely referred to as taxonomic importance are described. In so doing, perhaps methods that work better than NISP and MNI in some situations can be identified. And, we can explore how and why some of these methods are less accurate, valid, or reliable than NISP, MNI, or each other, and whether or not they should be used at all. This last point is critical because virtually all of the alternative methods discussed here have occasionally been advocated as better than NISP or MNI as measures of taxonomic abundances within a biocoenose. Because most of them were proposed 20 or more years ago, it seems appropriate to evaluate them in light of the new knowledge that has been gained over the past two decades.

The problems that attend NISP and MNI suggest that counting units different than MNI and NISP *should* be designed and used. And the literature contains arguments that counting units other than NISP and MNI should be used to determine taxonomic abundances. Clason (1972:141), for example, argues that MNI should be termed the “estimated minimum number of individuals,” and he uses the word *estimated* “intentionally because a real calculation of the minimum number of individuals is not possible.” He does not explain what he means, but given his other remarks it seems that he is concerned that MNI produces a *minimum* minimum. This is so because matching of bilaterally paired bones (left and right humeri, for example) will be less than perfect (some matches will not be identified, other matches will be incorrect), the true minimum number of individuals (MNI) (given that we cannot match each humerus with each tibia, each femur with each m<sub>3</sub>, etc.) represented by a collection will never be known. Of course that is true, but it also is fatalistic. Perfect data are seldom available in many scientific disciplines. Its absence from paleozoology is hardly a reason to not try to learn the limitations (analytical and interpretive) of

the data that are available. For example, do NISP values provide accurate ordinal scale measures of taxonomic abundances in a thanatocoenose or biocoenose? If so, then MNI values are unnecessary.

Another reason that alternative quantitative methods and units have been proposed as replacements for NISP and MNI is that the alternatives occasionally are designed to answer a different question than “Is taxon A more abundant than taxon B, and if so, by how much?” As I emphasized at the end of Chapter 2, explicitly defining the target variable that we are trying to measure should help us evaluate old measures and design better new ones. That is, in some cases, exactly what those who proposed the alternative measures discussed below had in mind. In the following, several of those alternative measurement units are reviewed. The first is one that is frequently advocated by zooarchaeologists and a related measure occasionally used by paleontologists – meat weight and biomass, respectively – that often rests on a calculation of MNI. The second is a quantitative method – ubiquity – that has seldom been used in paleozoology. Advocates of the third method – calculation of the Lincoln–Petersen index – argue that it provides a more accurate estimate of taxonomic abundances within the thanatocoenose or biocoenose than do either NISP or standard Whitean MNI values. Brief discussion of several other suggestions that have been made with respect to estimating the most probable number of individuals represented in a collection concludes the chapter.

#### BIOMASS AND MEAT WEIGHT

Paleontologists sometimes measure biomass, defined by biologists as the total amount of all biological tissue in a specified area or of a specified population. Paleontologists generally modify this definition to mean the total amount of biological tissue represented by taxa represented in the collection of animal remains they are studying (e.g., Damuth 1982; Guthrie 1968; Scott 1982; Staff et al. 1985). Zooarchaeologists (and sometimes ornithologists) also sometimes measure the amount of “meat” (or perhaps more accurately, the amount of consumable soft tissue) represented in a faunal collection (White 1953a). The amount of meat is some fraction of the biomass represented by an assemblage because not all tissue making up biomass is consumable. There are several methods that have been designed to measure biomass and several other ones designed to measure meat weight. Meat weight is usually a derivative of biomass, so we begin with biomass. It tends to be the less derived of the two given the methods used to calculate it.

### Measuring Biomass

In an early use of biomass, paleontologist R. D. Guthrie (1968:351) multiplied the “approximate annual average [live weight] of all age classes” of each species represented by their percentage frequency in each of four assemblages. He used genetically closely related modern taxa as an analog for the weight of individuals among the extinct prehistoric taxa represented in the assemblages. He used “annual” averages because of the marked seasonal changes in body weight among the mammalian taxa he was studying (e.g., Guthrie 1982, 1984a, 1984b). The average weight of all age classes accounted for the fact that youngsters of all taxa (plants and animals) weigh less than adults. Guthrie was particularly interested in the mammalian biomass that could be supported by what has subsequently been referred to as the “mammoth steppe” habitats of the late Pleistocene Arctic, so MNI was not the best measure of taxonomic abundances given the nuances of the question he was asking. Guthrie apparently used the equivalent of MNImin as the value for taxonomic frequencies in each assemblage, and multiplied that amount by the annual average live weight of all age classes to obtain his measures of biomass. Given that his research question concerned floral habitats, in his analysis he retained a distinction between grazers (indicative of grassland) and nongrazers.

Guthrie’s analysis is instructive for the simple reason that he was explicitly aware of several of the most significant variables that had to be dealt with were he to measure the prehistoric biomass of mammals. These include seasonal variations in body weight and ontogenetic (development or age) variation in body weight. The deer (*Odocoileus* sp.) and wapiti (*Cervus* sp.) remains from Cathlapotle illustrate the problems attending these variables. The MNI values, average live weights, and estimated biomass for each taxon in each assemblage are given in Table 3.1. The average annual live weight of all ages and both sexes reported by White (1953a) was used to estimate the biomass of the two taxa. The MNI abundances indicate deer outnumber wapiti slightly in both assemblages; the ratio of deer to wapiti based on MNI is 1:0.86 in the precontact assemblage and 1:0.89 in the postcontact assemblage. But as White (1953a, 1953b) likely would have predicted, given differences in body size, the biomass of wapiti – the larger of the two ungulates – is greater than that of deer in both assemblages; the ratio of deer to wapiti biomass is 1:3.4 in the precontact assemblage and 1:3.6 in the postcontact assemblage. Given that deer tend to browse a bit more than wapiti, and wapiti graze a bit more than deer, one might be tempted to conclude there was more grassland than shrubland or forest in the area. The biomass measures also suggest wapiti tissue is more abundant than deer tissue (in

Table 3.1. *Biomass of deer and wapiti at Cathlapotle*

	Average live weight (kg)	Precontact MNI	Precontact biomass	Postcontact MNI	Postcontact biomass
Wapiti	350	12	4,200	24	8,400
Deer	87	14	1,218	27	2,349

these collections), which contradicts the measure of taxonomic abundance provided by MNI.

Biomass estimates of the sort represented in Table 3.1 are not without problems. Most obviously, the biomass of deer and the biomass of wapiti are likely to not be ratio scale measures given that they are based on MNI, which is a measure that is likely to be, at best, only ordinal scale. Furthermore, if a method like that represented in Table 3.1 is used to calculate biomass, then the influence on the result of MNImin versus MNImax can be substantial. Aggregation with respect to calculating MNI plagues this kind of biomass measurement. And, there are other problems as well.

#### *Problems with Measuring Biomass (based on MNI)*

A variable that Guthrie did not worry about was whether his assemblages of faunal remains were coarse-grained palimpsests or fine-grained snapshots. He noted that each of the four assemblages he studied had been collected from deposits that represented “a relatively short duration” of time and that this “narrow unit of time permits the [paleoecologist] to consider these fossil assemblages as remnants of a community which occupied the immediate area where they were recovered” (Guthrie 1968:347). Exactly how much time was represented by the accumulation and deposition of each assemblage was unclear. That several hundreds or perhaps even thousands of years are represented by the assemblages would not be an unreasonable guess. Whatever the case, the point here is that paleozoological estimates of biomass are not measures of “standing crop” or the amount of biomass at *one point in time* (Krantz 1977).

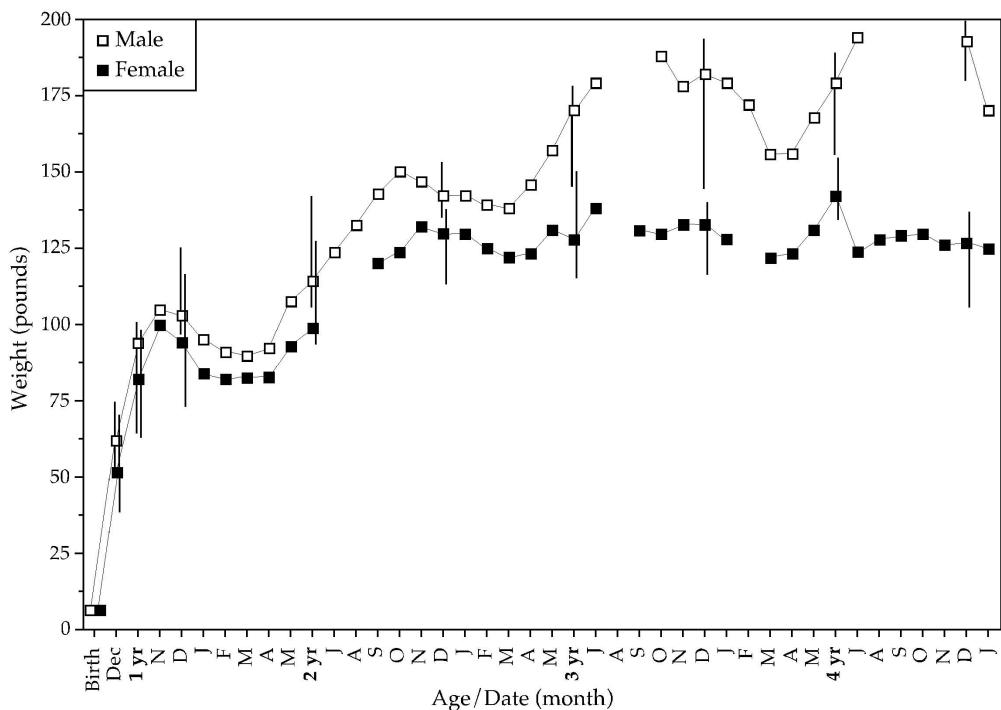
Biologists measure biomass at, effectively, one point in time; it may take them a month or two to actually measure it, but relatively speaking their measurements are fine grained. A paleozoologist, however he turns a collection of bones into a measure of biomass, is not measuring the same variable in terms of time that a biologist is. The paleozoologist is measuring that variable in terms of paleoecological time. Any collection of faunal remains with several taxa and several individuals of each

is likely to have been accumulated and deposited over some period of time greater than a year or even a decade. At present, we lack the taphonomic knowledge and paleochronometers that would allow us to determine the temporal duration over which a collection of faunal remains was accumulated and deposited. Such “time-averaged” collections may not always be a bad thing (e.g., Kowalewski et al. 1998; Lyman 2003b), but whether they are or not will depend on the temporal resolution required by one’s research question.

Another variable that Guthrie did not consider was individual variation; Guthrie used an average weight. All members of a taxon, even though they might be the same sex, the same age, the same health status, and are raised in the same habitat, will not weigh exactly the same as each other all of the time. Add in sex differences, age differences, health differences, minor habitat variation, and the range of individual variation increases accordingly. Figure 3.1 is redrawn from Brown (1961), a biologist who weighed live black-tailed deer (*Odocoileus hemionus columbianus*) in western Washington State. The figure shows a couple things relevant to measuring the biomass of deer, even assuming that we can derive MNI accurately. First, the two lines are each based on a single individual (one male, one female); minimums and maximums reported by Brown (1961) for other individuals for a limited number of months are also plotted in Figure 3.1. The potential magnitude of individual variation is remarkable.

Second, males are consistently larger than females, except perhaps at birth. Using MNI without distinguishing males and females masks sexual variation. The third thing to note in Figure 3.1 is the growth of deer over the first 2–3 years of life. They increase in size (and the biomass of individual deer increases) as they grow. Use of an average adult live weight as a multiplier ignores ontogenetic variation. Fourth, note the variation in weight by season in adult deer ( $\geq 3$  years of age). This is a particularly pernicious problem in temperate latitudes where seasonal variation in forage causes many animal species to gain weight in the spring, summer, and early fall, and lose weight in the late fall and winter (Guthrie 1984a).

But, you might suggest, we can control for ontogenetic age at death and season of death of the organism based on tooth eruption and wear, and we can control for sexual dimorphism by determining the sex of the individual represented by various animal remains. Certainly we can determine the age at death and the season of death of at least some deer and some wapiti represented in a collection based on tooth eruption and wear. In many instances with the deer and wapiti remains from Meier and from Cathlapotle, however, estimates of ontogenetic age were coarse – each estimate of age at death included a range of  $\pm$ several months – particularly for individuals older than about 4 years. Regardless of the age estimation process,



**FIGURE 3.1.** Ontogenetic, seasonal, and sexual variation in live weight of one male and one female Columbian black-tailed deer. Points are for one individual, vertical lines through points indicate observed range across multiple individuals. Modified from Brown (1961).

in no case were multiple lower teeth – the skeletal elements used to determine the age of deer and wapiti – the most common element and thus in no case did teeth provide both ontogenetic data and the MNI (Table 2.11). There is no way, then, to determine how many 6-month-old deer and wapiti, how many 12-month-old deer and wapiti, how many 18-month-old deer and wapiti, and so on, are represented without inadvertently omitting some of the forty-one deer (MNI<sub>max</sub>) and thirty-six wapiti (MNI<sub>max</sub>) represented (Table 3.1). Examine Figure 3.1 again and consider how such data might be used paleozoologically. Were the same kind of data available for wapiti, a pair of curves not unlike those in Figure 3.1 would likely be produced.

### *Solving Some Problems in Biomass Measurement*

Significant problems with Guthrie's protocol involve his use of an average adult live weight and simple MNI measures (no accounting for age, sex, or size variation) of

taxonomic abundance. More recently, paleobiologists who have estimated prehistoric biomass have distinguished taxa with high abundance in the fossil record from those with low abundance, and also distinguished taxa that included mostly large individuals from those that included mostly small individuals (Bambach 1993). A relatively abundant taxon made up of large individuals represents more biomass than a relatively rare taxon made up of small individuals. This analytical procedure solves many problems because it avoids measurement error. It does so by sacrificing resolution; taxonomic abundances are ordinal scale estimates as are estimates of modal body size of individuals in a population of a taxon. Biomass estimates based on these variables cannot be better than ordinal scale. Other forms of corrections have been proposed in the context of estimating meat weight, the measurement we turn to next.

### *Measuring Meat Weight*

White's (1953a) seminal procedure for measuring the amount of meat or consumable tissue represented by a collection of archaeological faunal remains was similar to Guthrie's for estimating biomass, plus one additional analytical step. First, determine the MNI (per taxon). Second, multiply the MNI (for a taxon) by the amount of meat one average individual (of the taxon) would provide. Third, an additional step (actually performed second in order according to White), involves multiplying the total live weight per average individual by the proportion of that weight thought to be edible. The mathematical equivalent would be to calculate the biomass of each taxon as Guthrie (1968) did, and then multiply that value by the proportion thought to be edible (White 1953a). In this case, for example, the biomass of deer and that of wapiti at Cathlapotle would, according to White (1953a), each be multiplied by 0.5 to obtain the amount of edible tissue each taxon provided. White provided "pounds of usable meat" for nearly three dozen species of mammals and more than two dozen species of birds.

White's procedure shares features with Guthrie's. For example, the biomass of a single individual (what he called "average live weight") that he used to derive usable meat amounts is an average of both sexes and all age classes, except for several sexually dimorphic species. Smith (1975) suggested that because many animal taxa are sexually dimorphic, such as deer and wapiti (but not considered as such by White), the analyst should establish the sex ratio in a collection and add that to the procedure for estimation of biomass. If some individuals were males and some were females, then step 2 of White's protocol had to be performed twice for a taxon, once for each sex. For example, in the postcontact assemblage at Cathlapotle, the sex ratio of wapiti

Table 3.2. *Meat weight for deer and wapiti at Cathlapotle, postcontact assemblage*

	Average live weight (kg)	Postcontact MNI	Postcontact biomass	Percent edible	Meat weight
Wapiti male	400	10	4,000	50	2,000
Wapiti female	300	14	4,200	50	2,100
Deer male	100	8	800	50	400
Deer female	70	19	1,330	50	665

is three males to four females, and the sex ratio for deer is three males to seven females (both based on anatomical features of the innominate). If those sex ratios are used to estimate the number of males and females among the total MNI, and then biomass and meat weight are recalculated using the estimated number of males and the estimated number of females (based on the observed sex ratio), the calculated values are different than when sexual dimorphism is not considered (Table 3.2). The ratio of deer to wapiti biomass increases from 1:3.6 without sexual dimorphism included to 1:3.8 when it is included. Because White's conversion factor to derive meat weight is 50 percent for both taxa and both sexes, the deer to wapiti meat weight ratio is also 1:3.8. We have gained resolution as to taxonomic abundances only if MNI values are ratio scale values *and* only if the estimated sex ratio is accurate. The latter may not be given that the total number of specimens that could be sexed is much less than twenty-four for wapiti and much less than twenty-seven for deer.

Smith (1975) also suggested that the analyst should determine the age structure or demography of each taxon represented by the collection. This could be done using species specific schedules of mandibular tooth eruption and wear. The proportion of the total MNI represented by each age class in the collection is then be added into the conversion of MNI into biomass. Estimating the number of individual deer and wapiti in each of several age categories at Cathlapotle is not possible. To get a feel for what is involved here, consider yet again Figure 3.1, and think about the fact that similar data, although not as fine-grained, exist for wapiti (Hudson et al. 2002).

Problems that attend White's (1953a) measurement protocol include the fact that the values for converting the biomass of one individual into mass of edible tissue are debatable (Stewart and Stahl 1977). White (1953a:397) derived his conversion values from (1) the gross morphology of the taxon (heavy-bodied, short-legged taxa vs. light-bodied, long-legged); (2) the percentage of live weight that professional butchers and meat packers estimated was usable meat; and (3) the presumed level of efficiency of primitive butchers. He thought his percentages would "give reasonably accurate

Table 3.3. Comparison of White's (1953a) conversion values (percentage of live weight) to derive usable meat with Stewart and Stahl's (1977) conversion values (percentages of live weight) to derive usable meat

Taxon	White	Stewart and Stahl
Mole ( <i>Condylura cristata</i> )	70	58–74 <sup>a</sup>
Rabbit ( <i>Oryctolagus cuniculus</i> )	50	40.7
Chipmunk ( <i>Tamias striatus</i> )	70	39.7
Squirrel ( <i>Sciurus carolinensis</i> )	70	26.0
Beaver ( <i>Castor canadensis</i> )	70	32.2
Muskrat ( <i>Ondatra zibethicus</i> )	70	51.9
Dog ( <i>Canis familiaris</i> )	50	80.8
Black bear ( <i>Ursus americanus</i> )	70	64.8
Fisher ( <i>Martes pennanti</i> )	70	64.8
Lynx ( <i>Lynx</i> sp.)	50	42.5
Seal ( <i>Phoca hispida</i> )	70	30
Deer ( <i>Odocoileus</i> sp.)	50	57 <sup>b</sup>

<sup>a</sup> based on two individuals.

<sup>b</sup> from Smith (1975).

results" and that any error would be "relatively constant," likely believing that errors were randomly distributed within and between taxa. Smith (1975) suggested that White's conversion factor for deer of 50 percent was too low and opted for 57 percent. Stewart and Stahl (1977:267) measured the body weight of a dozen carcasses of various taxa, then weighed the "edible tissues and organs," and calculated the latter as a percentage of the former. Overall, their values for the percentage of edible tissue of a carcass differ from White's (Table 3.3). Stewart and Stahl indicated that their data did not conclusively provide a set of conversion values that gave more accurate measures of edible tissue than White's. They hoped such an alternative set of conversion values could be developed, but no other set of conversion values has been proposed. Recent researchers have used White's (1953a) original values (e.g., Stiner 2005).

The problem identified by Stewart and Stahl (1977) may be greater than they imagined. It is likely that no set of conversion values that are consistently ordinal scale values, let alone ratio scale values, can be developed. Based on modern butchering practices of rifle-equipped hunters, data have been compiled on the amount of tissue that might be consumed. The percentage of live weight that comprises usable tissue of male and female wapiti of different ages varies considerably (Table 3.4). Were one to develop a set of conversion values such as Stewart and Stahl (1977) proposed, that set would need to include not just a conversion value for every species, it would need

*Table 3.4. Variation by age and sex of wapiti butchered weight (eviscerated, skinned, lower legs removed) as a percentage of live weight. From Hudson et al. (2002:250)*

Age	Male	Female
2 years	45	48
2 to 4 years	44	46
≥ 5 years	42	51

multiple conversion values for each species. A different conversion value would be necessary for each age class of each sex (Table 3.4). Figure 3.1 makes it clear that numerous conversion values are required for each taxon if one hopes to have ratio scale measures of usable meat. And this is not the only significant methodological hurdle to perfecting White's protocol.

Table 3.5 demonstrates that depending on how a carcass was butchered, and what the butcher thought was edible, the conversion values may vary considerably. For a mature male wapiti with a live weight of 350 kilograms, the proportion of live weight that comprises usable tissue varies more than 40 percent over several different stages of butchering. Not only that, whatever conversion value is used, that value assumes complete consumption – from nose through tail, as one anonymous commentator put it – of the carcass. As Binford (1978) demonstrated in an ethnoarchaeological context and Lyman (1979) argued from the perspective of historic zooarchaeological data, such an assumption is unwarranted (see also Schulz and Gust 1983). A single deer femur in an archaeological site does not necessarily represent the meat of an entire animal, regardless of the conversion value one uses to transform that bone specimen into amount of usable meat (or biomass). Lyman (1979) suggested determining the minimum number of butchering units, say, hindquarters of a species of mammal, and applying a conversion value to that quantitative unit. This simply shifts the problem of deriving a minimum number of carcasses – the standard MNI – to deriving a minimum number of each of several distinct butchering units (assuming such can be identified archaeologically), and it retains the problem of developing conversion values for multiple age–sex categories that may not be visible paleozoologically. The same problems attend similar suggestions by Schulz and Gust (1983) who used historical documents to estimate the rank order economic value of different butchering units (see also Huelsbeck 1989; Lyman 1987b).

Along these lines, Betts (2000:30) suggested that in an historic archaeological context the zooarchaeologist would do well to base estimates of meat amounts “on consumer units with a known relationship to the [faunal] material being analyzed.” He suggested that rather than calculate meat weight per taxon based on the MNI

Table 3.5. *Weight of a 350-kilogram male wapiti in various stages of butchering. From Hudson et al. (2002:250)*

Butchering stage	Weight of carcass	Proportion of live weight
Live weight	350	1.0
Bled weight	340	0.97
Eviscerated	228	0.65
Hide and lower legs removed	189	0.54

per taxon, the analyst should determine the frequency of consumer units per taxon. Consumer units might be standard quarters, wholesale units, or retail-like units of an animal, or something else. This procedure merely relocates the problem from converting an MNI value to meat weight to converting a (minimum?) number of consumer units to meat weight. Betts (2000) uses a conversion procedure not unlike White's (1953a), complete with all its attendant problems and weaknesses. Thus Betts (2000:30) correctly emphasizes that his suggested procedure "merely provides estimates of the actual meat contributions indicated by the remains." Those contribution amounts are at best ordinal scale.

We are left with a grim picture of measurements of meat weight using protocols such as Guthrie's and White's. Are there better techniques for measuring taxonomic abundances based on biomass or meat weight or both? Indeed, there are other ones, though these too have some serious weaknesses that make them rather tenuous.

### *The Weight Method (Skeletal Mass Allometry)*

A variable that one might measure is the weight of skeletal material per taxon. This quantitative unit was suggested by several zooarchaeologists in the 1960s and 1970s as one that could be used to measure either biomass or edible meat (Reed 1963; Uerpman 1973; Ziegler 1973). Despite significant criticisms (Casteel 1978; Chaplin 1971; Jackson 1989; Lyman 1979), zooarchaeologists the world over continue to measure and interpret the weight of osseous material (e.g., Dechert 1995; van Es 1995; Jackson and Scott 2002; Landon 1996; McClure 2004; Prummel 2003; Tuma 2004; Weinstock 1995). One zooarchaeologist has provided a detailed review of the nuances of this method, and argues that it should be studied further and perfected because of its value (Barrett 1993). It is worthwhile, then, to consider this quantitative variable in some detail.

First, the bone weight or, simply, weight method is based on the biological property of allometry. Allometry concerns the relationship of the size of one property of a body to the size of another. The size of a particular organ to the size of another in a body, the size of the head relative to the rest of the body, or the size of a limb relative to the size of the body are all allometric relationships. Allometry concerns the study of such relationships and if and how those relationships change during the ontogeny of an organism. The weight method developed by zooarchaeologists cashes in on the allometric relationship between bone weight and the total body weight of an organism. As biologists have noted, “animal skeletons scale allometrically with body mass, so that skeletons of large animals are proportionately more massive than those of small animals” (Prange et al. 1979:103). The ratio of bone weight to live weight per individual is greater for large individuals (whether of the same or different taxa) than that ratio is for small individuals (Needs-Howarth 1995). The presumption is that if the statistical nature of the relationship between these two variables can be determined, then that relationship can be used to convert bone weight observed in an archaeological collection into biomass or usable meat weight.

As summarized by Casteel (1978), many of the original analytical protocols for using the weight method involved two steps once the specimens comprising a collection of faunal remains had been identified to taxon. First, weigh all remains of each taxon separately. Then, convert the bone weight of each taxon to either a measure of biomass or of usable meat weight. As might be expected given the comments on White’s (1953a) conversion values presented earlier, the conversion values proposed by those advocating the weight method varied considerably from author to author. Cook and Treganza (1950:245), for example, estimated that dry bone represented about 6 percent of the original live weight of a mammal and of a bird. Adopting this estimate, the weight of dry bone would be converted to biomass by multiplying that weight by 16.67 (Casteel 1978). In contrast, Reed (1963) estimated that dry bone weight comprised about 7.5 percent of the original live weight of a mammal. Were one to adopt this estimate, the weight of dry bone would be converted to biomass by multiplying that weight by 13.33 (Casteel 1978). As Casteel (1978) pointed out, empirical studies (as opposed to Cook and Treganza’s and Reed’s estimates) suggested that anywhere from about 8.5 to 13 percent of the body weight of mammals constituted bone weight. Ignoring the slippery issue of choosing which percentage to use, once you have biomass per taxon, you may want to convert that to usable meat, which introduces problems like those associated with White’s (1953a) analytical protocol.

Shortly after the weight method began to see some use, Chaplin (1971:68) noted that anyone using it had to assume that the relationship between bone weight and biomass (or usable meat weight, whichever was sought) was constant. This was so because

the conversion value chosen was a constant; it did not vary. Thus, if 5 kilograms of bone represented 60 kilograms of biomass, then 10 kilograms of bone represented 120 kilograms of biomass, 15 kilograms of bone represented 180 kilograms of biomass, and so on. In Chaplin's eyes, such an assumption was not warranted. Furthermore, in some ways anticipating later criticisms of White's (1953a) MNI-based method, Chaplin (1971) noted that there was some controversy over which conversion value to use given individual variation within a taxon as well as the fact that an individual's weight would vary over time (recall Figure 3.1). Chaplin (1971:68, 69) concluded that the "relationship of bone weight and body weight is not an exact one" and he recommended "the meat/bone ratio must be established for each bone, at different ages of the animal and for each sex." The meat—bone ratio constitutes recognition that people might not have consumed an entire carcass, that 5 kilograms of phalanges do not represent the same amount of biomass and usable meat as do 5 kilograms of femora of the same taxon, and that 5 kilograms of humeri from 6-month-old females do not represent the same amount of biomass as do 5 kilograms of tibiae from 36-month-old males of the same taxon.

Those who have used bone weight as a measure of taxonomic abundance over the last 15–20 years usually give reasons as to why it is a measure that *should* be used. They also often state that it is a better measure of taxonomic abundance than either NISP or MNI. Those arguments can be summarized as follows:

- 1 Measures of bone weight allow the summation or merger of abundance data into more general taxonomic categories such as a taxonomic family or "large mammal."
- 2 Measures of bone weight are not influenced by fragmentation.
- 3 Measures of bone weight circumvent interanalyst variation in identification skills that bias measures of taxonomic abundance.
- 4 Because of the statistical relationship between bone weight and body weight, bone weight provides proxy measures of usable meat and of the importance or contribution of a taxon to diet.

This list may seem impressive if not overly long. Critical scrutiny of the arguments indicates, however, that each is a justification to not use other measures of taxonomic abundance (particularly NISP and MNI) rather than a warrant to use bone weight. This is so because all four statements are easily shown to be variously false, of minimal significance, or to apply to bone weight as well as to other measures of taxonomic abundance.

The first reason given to use bone weight is certainly true because one can sum the weights of specimens identified as, say, deer with the weight of specimens identified

as deer size. The significance of the first statement resides, however, in the unspoken underpinning assumption that were one to use NISP or MNI to quantify taxonomic abundances, one would not determine the MNI for categories like a taxonomic family or large mammal or deer size. That is false; these same categories have been used by paleozoologists. A well-known example is the five size categories of African bovid that zooarchaeologists use because of difficulties with identifying the genus or species of bovid represented by bones and teeth (Brain 1981; Bunn and Kroll 1986).

The second reason given to use bone weight is also true, but it ignores the fact that increasingly intensive fragmentation makes identification progressively more difficult. A fragment of medium- or small-size mammal long bone diaphysis might be confused with a similar fragment from a large- to medium-size bird, for example (Driver 1992). Thus fragmentation can influence which bones are weighed – those that are identifiable are weighed, and intensively fragmented specimens will be unidentifiable. The second reason also relates to the first in the sense that fragments that cannot be identified to species but that can be identified to, say, taxonomic family can indeed be weighed (just as they can be tallied for the NISP of a taxonomic family), but at the cost of losing fine-scale taxonomic resolution (just as when fragments identifiable to genus or species are included in NISP tallies of a taxonomic family).

The third reason given to measure bone weight identifies a real problem regardless of whether one quantifies taxonomic abundances with NISP, MNI, or bone weight (Gobalet 2001; Lyman 2005a). Specimens that are not identified – regardless of whether the categories of identification are species, genera, families, large, medium, and small mammal, or whatever – are simply not tallied nor are they weighed. Therefore, this reason, like the preceding two, is not a valid warrant to measure bone weight.

The fourth reason contends with the fact that neither an NISP of one mouse and of one elephant, nor an MNI of one mouse and of one elephant provides an accurate measure of the contribution of those two taxa to diet. But the fourth reason is not a good warrant to use bone weight to get at usable meat weight. The problem was originally identified by Chaplin (1971). If one converts bone weight to biomass or weight of usable meat, one must assume the relationship between the two variables is constant and linear. That is, one assumes a single conversion value such as 7.5 percent of an (average?) individual's total weight represents the weight of the skeleton, and the remainder is soft tissue. Chaplin (1971) argued that the relationship is in fact not constant, and Casteel (1978) showed that the relationship is neither constant nor linear.

Casteel (1978) used previously published data (McMeekan 1940) on the relationship between bone weight and biomass of a domestic pig (*Sus scrofa*) to show that the relationship between the two variables is curvilinear. Casteel (1978) examined

Table 3.6. Descriptive data on animal age (weeks), bone weight per individual, and soft-tissue weight per individual domestic pig. All weights are grams. From McMeekan (1940)

Age	Bone weight	Muscle weight	Muscle + fat weight	Muscle + fat + hide weight
Birth	242.7	388	439	545
4	767	1,901	2,885	3,240
8	1,730	4,182	6,156	6,990
16	3,962	12,669	19,796	21,534
20	5,214	17,718	30,904	33,198
24	6,438	23,015	43,904	46,979
28	7,396	31,647	66,160	69,602

the relationship between the bone weight of a single individual pig and the weight of muscle tissue of that individual, and the relationship between the bone weight of a single individual and the total soft-tissue weight of an individual (muscle + fat + hide) because he was unsure how completely the soft tissues of a pig might be used by prehistoric consumers. He found that the relationships between both variable pairs varied with the ontogenetic age of the pig; the ratio of bone weight to soft-tissue weight increased as ontogenetic age increased.

I used the same data (Table 3.6) and expanded Casteel's analysis to include another variable pair – the relationship of an individual's bone weight and that individual's muscle + fat tissue weight. I replicated Casteel's results for the two variable pairs he examined, although my statistics are slightly different than his, likely as a result of my use of a different computer program (although the negative sign for the calculated Y intercept is not included in the published version of Casteel's analysis). The statistical relationships between bone weight per individual animal and the weight of various categories of soft tissue are summarized in Table 3.7 and illustrated in Figure 3.2. The relationship between bone weight and muscle weight, between bone weight and soft-tissue weight, and between bone weight and soft tissue, regardless of the soft tissues included, are all curvilinear. The relationship is described by the power-function formula  $Y = aX^b$ , where  $X$  is the bone weight per individual,  $Y$  is the soft-tissue (however defined) or complete carcass weight,  $a$  is the Y intercept,  $b$  is the slope of the best-fit regression line, and  $a$  and  $b$  are empirically determined. Regardless of how soft tissue is defined for the domestic pig data, the coefficient of determination ( $r^2$ ) is greater than 0.99; more than 99 percent of the variation in soft-tissue weight is accounted for by variation in bone weight (Table 3.7).

Table 3.7. Statistical summary of the relationship between bone weight ( $X$ ) and weight of various categories of soft-tissue ( $Y$ ) for domestic pig (based on data in Table 3.6)

Soft tissue	Regression equation	$r$	$p$
Muscle <sup>a</sup>	$Y = -0.38(X)^{1.25}$	0.999	< 0.0001
Muscle + fat <sup>a</sup>	$Y = -0.47(X)^{1.35}$	0.997	< 0.0001
Muscle + fat + hide	$Y = -0.66(X)^{1.39}$	0.996	< 0.0001

<sup>a</sup> also determined by Casteel (1978).

Table 3.7 and Figure 3.2 highlight the fact that the weight of the skeleton in an individual is tightly related to the weight of the soft tissue of that individual. But the relationship between the two variables is not constant; it is not linear. Thus were one to develop a means to estimate biomass or usable meat from the weight of bone, a single conversion factor would not provide accurate results. To produce accurate results, one should empirically derive an equation in the form of a power function

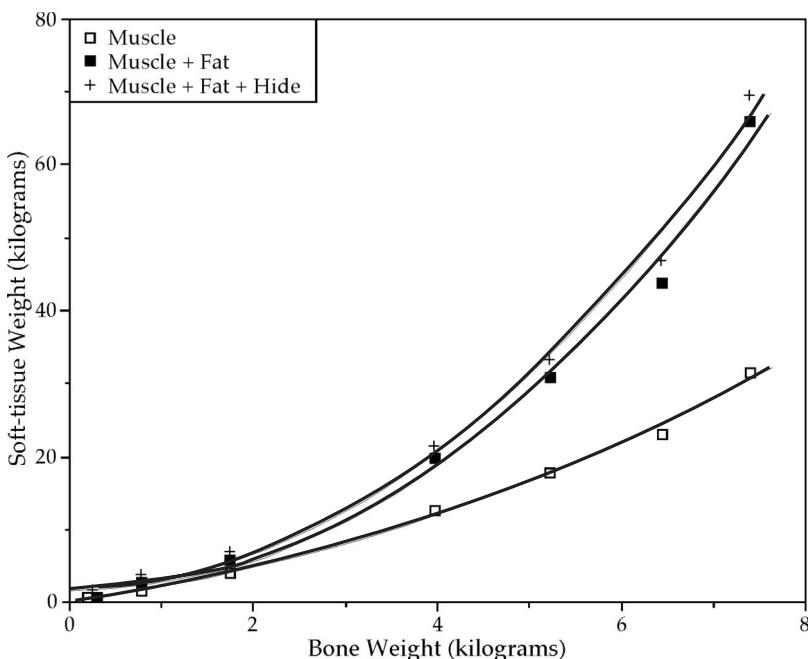


FIGURE 3.2. Relationship between bone weight per individual and soft-tissue weight in domestic pig. The curves are best-fit, second-order polynomials. Data from McMeekan (1940); see Tables 3.6 and 3.7.

to account for the curvilinear (allometric) relationship between bone weight and soft-tissue weight. But even with such an equation, various difficulties remain. For example, what conversion value should be used to transform a measure of biomass into a measure of usable meat, the actual variable zooarchaeologists who advocate the weight method want to measure (e.g., Barrett 1993)? A different formula for each taxon would control for intertaxonomic variation. What about intrataxonomic variation? And there are several other slippery issues as well.

Many commentators have pointed out that preservational (diagenetic or post-burial) conditions will alter the weight of bones, and thus the relationship between archaeological bone weight and soft-tissue weight will be altered (Barrett 1993; Lyman 1979; Wing and Brown 1979). Casteel (1978) noted that empirically deriving an equation that describes the relationship between meat weight and bone weight of an individual presumes that a collection of bones from a site represents a single (perhaps impossibly) large individual. This is another way of saying that the amount of meat associated with phalanges of an individual is not the same as the amount associated with the scapulae or the femora of an individual. Jackson (1989:604) put it this way: the weight method treats “all bone fragments of a given weight as if they supported a similar amount of tissue regardless of the element from which they originated. . . . [The] formulae treat bone weight *as if* it came from a single individual.” If a sample consists of more than a few specimens, it is likely that those specimens represent more than one individual. Thus Chaplin (1971) recommended that a conversion value be determined for each bone in the skeleton or for each portion of a skeleton. Is that possible?

Chaplin’s concern, as well as the concern of many others who have commented on the problem, is that the skeletal mass allometry equations used by zooarchaeologists do not account for the fact that a single skeleton of deer may weigh the same as, say, fifty deer metacarpals. But formulae such as those for domestic pigs (Table 3.7) presume that one has weighed one or more skeletons, not piles of metacarpals, or collections of selected skeletal elements from various individuals. Data on weights of carcass portions and bones published by Binford (1978) show that Chaplin’s suggested solution does not resolve all of the problems with the weight method. Binford’s data are for two domestic sheep (*Ovis aries*); they are for a 6-month-old lamb and a 90-month-old female (Table 3.8). The relationship between bone weight and gross weight for each carcass portion for each individual is graphed in Figure 3.3. The statistical relationships between the two sets of data are summarized in Table 3.9. The graph (Figure 3.3) indicates that the relationship is steeper in the older sheep (slope = 1.063) than in the younger individual (slope = 0.826); Casteel’s (1978) data for the pig might prompt us to make similar predictions. The ratio of bone weight to biomass is greater

Table 3.8. Descriptive data on dry bone weight per anatomical portion and total (soft-tissue + bone) weight per anatomical portion for domestic sheep. All weights are grams. Weights for limbs are for one side only. Data from Binford (1978:16)

Skeletal portion	6 month old, dry bone weight	6 month old, total weight	90 month old, dry bone weight	90 month old, total weight
Skull	152.40	317.52	294.82	938.05
Mandible	92.02	408.24	167.60	1,193.87
Atlas-axis	53.08	272.16	87.90	408.24
Cervical	73.40	725.76	137.40	1,088.64
Thoracic	91.60	637.27	288.58	1,758.20
Lumbar	69.68	315.29	205.35	871.29
Pelvis-sacrum	122.34	1,140.08	319.80	1,623.55
Ribs	121.90	1,360.80	373.04	1,995.84
Sternum	18.47	907.29	52.75	1,859.76
Scapula	29.50	556.80	75.10	844.76
Humerus	56.50	385.47	95.10	584.86
Radius-ulna	45.30	214.15	88.50	324.90
Metacarpal	32.10	86.81	51.50	135.08
Phalanges (front)	16.20	71.04	38.40	106.41
Femur	73.17	985.75	121.00	1,474.20
Tibia & tarsals	56.96	282.69	114.00	498.96
Metatarsal	51.50	140.61	59.49	149.69
Phalanges (rear)	16.20	55.65	38.40	99.79

in older than in younger individuals, apparently even across the skeleton. The point scatters also suggest that the relationship between bone weight and gross weight is not very tight regardless of the age of an individual; the coefficient of determination ( $r^2$ ) is  $< 0.6$  for both. If these data are representative of the relationship between bone weight and biomass, they are also representative of the relationship between bone weight and weight of soft tissue. They suggest that intrataxonomic variation, or individual variation in the relationship of the two variables, may be difficult to account for in any single formula for a taxon.

Consider how the variation shown in Figure 3.1 relates to that shown in Figure 3.3. We are seeing a reflection of some of the same (individual or intrataxonomic) sources of variation in both. It is precisely for this reason that Barrett (1993), the strongest recent advocate of perfecting and using the weight method, must conclude that the method at best produces ordinal scale data on biomass and usable meat, and may not

Table 3.9. Statistical summary of the relationship between bone weight ( $X$ ) and gross weight or biomass ( $Y$ ) of skeletal portions of two domestic sheep (based on data in Table 3.8)

Sheep age	Regression equation	$r$	$r^2$	$p$
6 months	$Y = 1.109X^{0.826}$	0.593	0.350	0.0095
90 months	$Y = 0.6X^{1.063}$	0.759	0.576	0.0003

even produce that scale of resolution. Barrett (1993:11) does not use these words, but instead indicates that the method is perhaps best used to estimate “a range of meat yield estimates for groups of excavated bone” and suggests that “meat yield estimates can be graphed [with] error bars to reveal broad patterns in the archaeological assemblage.” Thus he presents an exemplary range of meat yield values as follows: mammals – 345.31 to 441.5 kilograms; fish – 82.26 to 140.46 kilograms; and birds – 0.58 to 1.07 kilograms. Such data are clearly ordinal scale. In this case the ranges do not overlap, but the discussion in Chapter 2 implies that biomass and usable meat

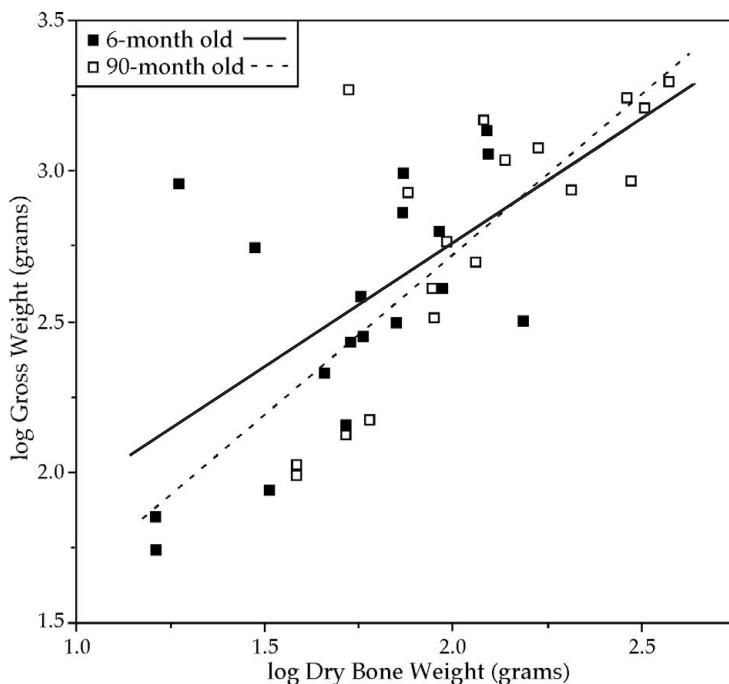


FIGURE 3.3. Relationship between bone weight per skeletal portion and gross weight per skeletal portion in 6-month-old domestic sheep and a 90-month-old domestic sheep. Simple best-fit regression lines are shown for reference. Data from Table 3.8.

or meat yield may not be even ordinal scale data. And based on discussions in this chapter and in Chapter 2, it is likely that ranges of biomass per taxon may overlap, making them nominal scale data.

### Bone Weight

Can we use the weight of skeletal tissue as a fundamental measure of taxonomic abundance? Uerpmann (1973) suggested using the weight method to derive meat weight, then using a conversion factor to change meat weight to numbers of individuals. A procedure that is mathematically the reverse of that proposed by White (1953a) is required to perform Uerpmann's second conversion, and thus is subject to all of the problems that attend White's analytical protocol. And, the first conversion must contend with all the problems with the weight method presented thus far. In light of these facts, it is no surprise that no one has actually done what Uerpmann suggested. What some individuals have done, however, is to use bone weight as a fundamental measure of taxonomic abundances (McClure 2004; Tuma 2004). That is, bone weight per taxon is recorded and then interpreted as is, without being converted to biomass or to usable meat. Does this procedure solve or circumvent problems that attend NISP and MNI as measures of taxonomic abundances? It is easy to show that bone weight per taxon has problems of its own.

First, as Barrett (1993:3) points out, this "method is attractively simple but it requires an assumption that the bone weight to body weight ratios for different [taxa] are virtually identical." Bone weight might indicate which taxon represents the most biomass (regardless of how much soft tissue is involved) but only if the ratio of bone weight to body weight is the same across all taxa. But we know that different taxa have different ratios (as do different individuals within each taxon). Thus, to use a simple example, if one taxon's average (to account for individual variation) ratio is 5 percent and another taxon's average ratio is 10 percent, then 10 kilograms of bone of each represent 200 kilograms of biomass for the first taxon and 100 kilograms of biomass for the second taxon. Bone weight indicates that the two taxa are equally abundant, but their biomass indicates that they are not.

Does bone weight provide information about taxonomic abundances in general that is not provided by, say, NISP? An immediate objection that this could not possibly be the case might involve noting that a bison (*Bison bison*) or a domestic cow (*Bos taurus*) has more or less the same number of skeletal elements as a squirrel (*Spermophilus* sp. or *Sciurus* sp.) or a mouse (*Microtus* sp. or *Mus* sp.), but the bones that comprise the skeleton of the former two taxa are much larger than the bones

of the skeletons of the latter two taxa. A femur of a cow will weigh considerably more than the femur of a mouse. And given the relationship between bone weight and biomass or weight of soft tissue, a mouse femur cannot represent the same biomass or the same amount of soft tissue as a cow femur. These observations do not comprise a reason to reject the possibility that bone weight might duplicate the taxonomic abundance data provided by NISP. Are NISP and bone weight related in such a manner as to be redundant measures of taxonomic abundances?

To answer the question just posed, I determined the correlation between bone weight and NISP of mammal remains in seventeen collections from North America, South America, and the Near East (Table 3.10). These collections represent seven to twenty-two taxa; some collections date to the historic period, others are prehistoric in age. They were chosen because I had access to published reports describing them. NISP and bone weight are significantly correlated ( $p \leq 0.05$ ) in fourteen of the seventeen collections. In so far as these seventeen collections are representative of all such data sets, they suggest that, with respect to variation in abundances of at least mammalian taxa, bone weight is often redundant with NISP as a measure of taxonomic abundance. It is beyond my scope here to explore in detail why these two variables are often correlated.

A vocal advocate of using bone weight allometry as a measure of taxonomic abundance has been Elizabeth Reitz (and colleagues) at the University of Georgia. Using curated modern skeletons with associated live weight data, Reitz developed several formulae of the power function form that allow one to convert archaeological bone weight of general taxonomic categories such as mammals, birds, turtles, and several categories of fish into biomass (Reitz and Cordier 1983; Reitz et al. 1987). These formulae were repeated and used many times during the 1980s and early 1990s (various references in Table 3.10), and were repeated yet again in a zooarchaeology text book (Reitz and Wing 1999) and in articles published in the third millennium (Reitz 2003). That a plethora of intrataxonomic and intertaxonomic variations were smoothed away by the use of a single formula in the general taxonomic category went largely unremarked.

When advocated, the allometric formulae were characterized as “based on samples drawn from known biological populations; the archaeological data [to which they were applied] are considered samples of archaeological populations rather than of individuals. This is the case even when original live weight is estimated for individuals” (Reitz et al. 1987:307). The key problem was also recognized: “Use of archaeological bone weight as the independent value merely predicts the amount of body mass that amount of bone could support *as if* the bone represented the skeletal mass of a real animal” (Reitz and Cordier 1983:247). If a deer skeleton weighs 25 kilograms,

Table 3.10. Relationship between NISP and bone weight of mammalian taxa in seventeen assemblages

Assemblage	N of taxa	r	p	Reference
Winslow	10	0.929	< 0.0001	Landon (1996)
Spencer-Pierce	10	0.791	= 0.0065	Landon (1996)
Paddy's	7	0.927	= 0.0027	Landon (1996)
Feature F	8	0.884	= 0.0035	Tuma (2004)
Feature E	9	0.920	= 0.0004	Tuma (2004)
Three sites	11	0.963	< 0.0001	McClure (2004)
Hirbet-Ez Zeraqon	10	0.942	< 0.0001	Dechert (1995)
Tell Abu Sarbut	20	0.922	< 0.0001	van Es (1995)
Carthago	11	0.903	< 0.0001	Weinstock (1995)
New Halos	7	0.659	= 0.108	Prummel (2003)
Washington, AR	7	0.884	= 0.008	Stewart-Abernathy and Ruff (1989)
SE Coast	14	0.896	< 0.0001	Reitz and Honerkamp (1983)
Paloma	8	0.863	= 0.0058	Reitz (1988)
NW Company	10	0.966	< 0.0001	Ewen (1986)
Pirincay	22	0.767	< 0.0001	Miller and Gill (1990)
Mose	14	0.530	= 0.0514	Reitz (1994)
Iroquois	8	0.651	= 0.08	Scott (2003)

then 25 kilograms of phalanges will suggest a biomass equivalent of one deer. Because the bone-weight allometric equation is founded on individual animals, it produces results that are not ratio scale measures of biomass when applied to archaeological collections consisting of a few bones from each of several skeletons, and they may not be ordinal scale.

Using the data in Table 3.8 five “collections” of 10 skeletal portions each were generated by randomly drawing individual skeletal portions and random assignment of each portion to either the 6-month-old sheep or the 90-month-old sheep of Binford (1978). Using the data in Table 3.8, the total bone weight and the gross weight of each was summed for each of the five collections. Beginning with one collection, another collection was successively added until five collections had been generated. This produced five collections of different sizes (number of skeletal portions varied). Finally, the general bone weight allometry formula for mammals generated by Reitz

Table 3.11. Results of applying the bone-weight allometry equation ( $Y = 1.12X^{0.9}$ ) of Reitz and Cordier (1983) to five collections of domestic sheep bone randomly generated from Table 3.8. All weights are grams

Collection	Bone weight	N of skeletal portions	Allometry biomass	Actual biomass (from Table 3.8)	Difference
1	837.07	10	5,623.4	6,169.8	546.4
2	2,103.93	20	12,882.5	14,454.8	1,572.3
3	3,279.75	30	19,054.6	22,560.8	3,506.2
4	4,008.04	40	22,908.7	27,357.9	4,449.2
5	4,810.76	50	26,915.3	31,938.8	5,023.5

and Cordier (1983) was used to estimate biomass for each of the five collections. That formula is:  $Y = 1.12X^{0.9}$  where  $Y$  is biomass,  $X$  is bone weight, 1.12 is the  $Y$  intercept, and 0.9 is the slope. (Another way to present this equation that makes it mathematically easier to calculate by hand is:  $\log Y = 1.12 + 0.9X$ .) Results of applying this formula to the randomly generated collections of sheep bone are shown in Table 3.11 where it is clear that the bone-weight allometry formula consistently underestimates biomass as measured by summing individual skeletal portions, and it does so with increasing magnitude as bone weight increases. This occurs because the bone weight allometry equation does not allow for intrataxonomic variation, nor does it account for the fact that the equation is built as if archaeological bone weight concerns complete skeletons, not various bones representing incomplete skeletons. And if that is not enough to worry about, aggregation will influence the results of applying the bone-weight allometry equations because different aggregates will distribute bones and thus bone weight differently across collections.

Biomass data determined by the skeletal mass allometry method are likely at best ordinal scale for many reasons. What has not previously been noted is that the taxonomic distribution of biomass data for faunal collections tends to look much like that distribution for NISP and MNI data (Figures 2.13–2.16). The distributions for two faunas described by Quitmyer and Reitz (2006) are shown in Figure 3.4. To generate these figures only data for vertebrate taxa (mammals, fish, birds, reptiles) identified to at least the genus level were used. The taxa represented by low biomass values tend to tie or differ minimally in value whereas taxa represented by high biomass values tend to differ in value by considerable amounts. As with NISP and MNI data (see Chapter 2), taxa with low biomass amounts are likely not even ordinal scale but instead nominal scale. Taxa with high biomass amounts may well be ordinal

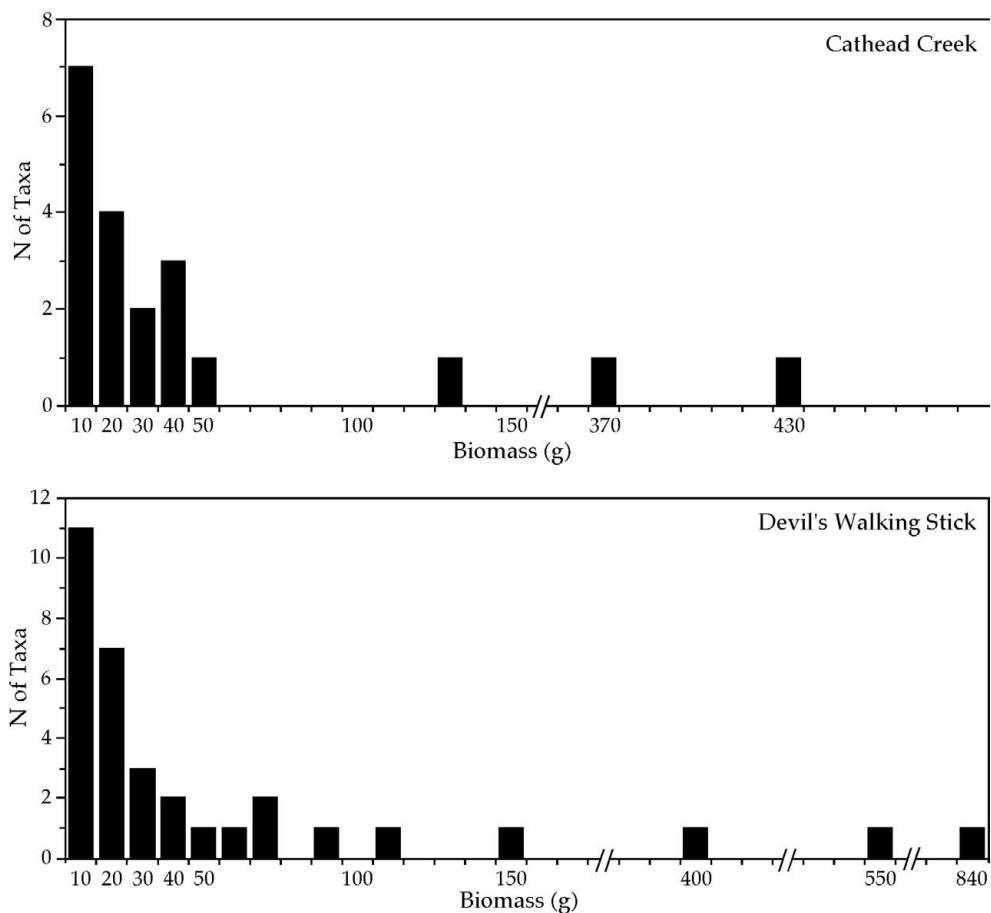


FIGURE 3.4. Frequency distributions of biomass per taxon in two sites (Cathead Creek and Devil's Walking Stick) in Florida State. Data from Quitmyer and Reitz (2006).

scale. Lest one think this distribution is a function of the multiple taxonomic groups included in Figure 3.4, Figure 3.5 shows a similar distribution for the biomass of mammals only in a collection described by Carder et al. (2004).

The most serious problems with bone weight allometry, then, are two. First, the method smoothes intrataxonomic and intertaxonomic variations. As Needs-Howarth (1995:94) correctly observed, “like average meat weight computations [of Guthrie and White], this method cannot take into consideration differences in [individual] condition.” Second, the bone weight allometry method is based on the weight of complete individual skeletons, yet is applied to commingled not-necessarily random accumulations of bones from multiple skeletons. Both of these problems are

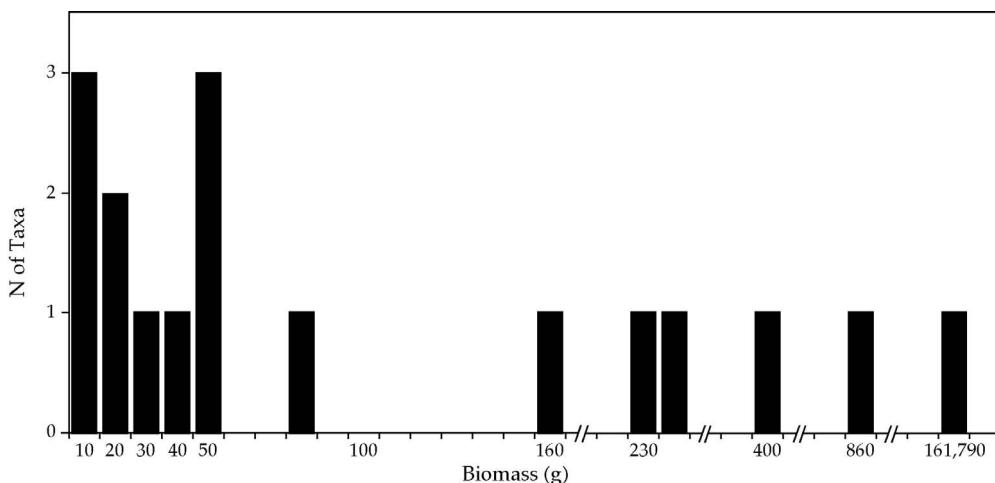


FIGURE 3.5. Frequency distributions of biomass per mammalian taxon in a site in Georgia State. Data from Carder et al. (2004).

dealt with in a subtle way by the method's advocates. Biomass amounts estimated using the allometry formulae are said to be “hypothetical amounts” (Reitz and Cordier 1983:247), to be “approximate” measures of abundance (Reitz and Cordier 1983:248), or to be “estimates” (Reitz et al. 1987:307). These are other ways to say that biomass estimates are, at best, ordinal scale.

The argument that bone weight allometry formulae should be used because they provide abundance data in a “morphological nutritional format” (Reitz and Cordier 1983:248) is weak. The bone weight allometry formulae provide at best ordinal scale data on taxonomic abundance that may be redundant with taxonomic abundance measures based on NISP. At least, that seems to be the case with the class Mammalia. It may not be the case when bone weights for other classes of vertebrates or invertebrates are added to the mix. But it is likely that allometry formulae for birds, finny fish, shellfish, reptiles, amphibians, and other categories of animals will be subject to the same kinds of problems as are described in Tables 3.8 and 3.11.

Advocates of the weight method seem to be aware of the fact that bone weight allometry, although based on biological and physiological principles, and although mathematically elegant, produces at best ordinal scale data. They indicate that “It seems unwise to generate allometric equations for each sex or for various [ontogenetic, seasonal] weight conditions, even if a data base could be acquired; to do so would limit archaeological applications” (Reitz et al. 1987:311). What is meant here is that sexual variation, ontogenetic variation, and seasonal variation can not

always be determined from archaeological remains, so to construct allometry formulae that account for these sorts of variation is unnecessary. It is, of course, precisely such variation that reduces mathematically elegant solutions, whether Guthrie's and White's simplistic average live weight, or Uerpmann's bone weight conversion, or allometric formulae, to producing at best ordinal scale results.

One variant of skeletal mass allometry is Needs-Howarth's (1995:95) suggestion that "once the soft-tissue biomass weight has been estimated, its caloric content can be calculated." She used skeletal mass allometry to estimate soft-tissue biomass, and then used U.S. Department of Agriculture standards to estimate calories per gram of edible tissue. She correctly noted that "edible tissue" likely varied cross-culturally, that individual variation in animal condition was masked by the procedure, and that the caloric results comprised "a relative [ordinal scale], indirect [derived], and probably distorted quantification of what the inhabitants of the site actually ate" (Needs-Howarth 1995:99). These comments suggest that the probability of distortion is  $> 0.99$ , and the degree of distortion may render the results nominal scale. Yet Needs-Howarth (1995:97) echoes those advocating skeletal mass allometry when she argues that soft tissue biomass estimates, "insofar as these are accurate for archaeological material, provide a more biologically justifiable estimate of potential food intake represented by excavated bone, than do NISP or MNI." Although true (largely because NISP and MNI are not meant to estimate "food intake"), the critical question concerns the accuracy of the soft-tissue biomass estimates, whether or not those estimates are converted to calories.

#### BONE SIZE AND ANIMAL SIZE ALLOMETRY

There is another technique for measuring biomass of taxa that is also based on allometry. This technique uses one or more linear measurements of bone size to estimate individual body size or biomass (e.g., Casteel 1974; Noddle 1973; Witt 1960). In at least one instance it was explicitly proposed as an alternative to White's (1953a) procedure for estimating meat weight (Emerson 1978, 1983). What is sometimes referred to as *linear allometry* also received some consideration by those who examined bone-weight allometry (Reitz and Cordier 1983; Reitz and Wing 1999; Reitz et al. 1987). The latter quickly dispensed with the bone size allometry option because it did not include all (weighable) bone and thus deleted some potentially informative data. Furthermore, it was noted that if the measured skeletal element or part was not also the most common element and thus was not used to define MNI, then any measure of biomass based on bone size would necessarily be an underestimate because a number

of specimens less than the MNI would provide the data. As well, the allometry formulae were based on, typically, one linear dimension yet were meant to estimate live weight of a complete animal. Bone weight advocates noted that such a procedure assumed an animal was eaten completely, from nose to tail, yet bone weight did not require that assumption. Finally, bone weight advocates implied that linear dimension allometry demanded species-level identification and thus could not incorporate into the analysis many specimens, yet bone weight allometry could incorporate those specimens identified only to taxonomic class or order. Nevertheless, linear allometry has seen some use in zooarchaeology. Before reviewing that use, consider how linear allometry is used in paleontology.

Vertebrate paleontologists have long used linear allometry to estimate the size of prehistoric animals (references in Damuth and MacFadden 1990). They are thus well aware of problems such as those identified by the advocates of skeletal mass allometry. Paleontological interest in body size is a direct result of the relationships between body size and functional anatomy, physiology, and metabolism, as well as the paleoecological implications of body mass inherent in, for example, Bergmann's rule (Blackburn et al. 1999). Thus, solving or circumventing problems of linear allometry methods have in the past 10–20 years become quite important in paleontology (e.g., Anyonge 1993; Anyonge and Roman 2006; Egi 2001; Mendoza et al. 2006; Reynolds 2002; Smith 2002; Wroe et al. 2003). Some of the problems with skeletal mass allometry are avoided by using multiple linear dimensions of multiple skeletal elements to estimate the sizes of bodies. Thus, femur length may suggest one body size but the breadth of the proximal humerus suggests another body size, which encounters head on the problem of possible skeletal element interdependence. Interdependence is, of course, not a problem if only one skeleton is involved. But the use of multiple dimensions and multiple elements avoids the problem of measured bones being fewer than the MNI; it also avoids the equivalence of 10 kilograms of phalanges giving the same amount of biomass as 10 kilograms of femora. And, measuring multiple bones, particularly if multiple dimensions of each kind of skeletal element are measured, provides more than a single data point (such as bone weight does) from which to estimate body mass.

Regression equations are built from known comparative skeletons and the independent variable (a dimension of bone size) is plugged in and the equation solved to determine a value of the dependent variable (body mass). A paleontologist may or may not attempt to estimate the total biomass represented by the bone collection, and does not often use the derived estimates of body size to estimate taxonomic abundances in the form of biomass. Because those analytical steps are seldom taken, the number of inferential layers, one built atop another, and the number of attendant

Table 3.12. *Deer astragalus length (millimeters) and live weight (kilograms). Data from Emerson (1978)*

Astragalus length	Weight	Astragalus length	Weight
26.0	7.53	41.0	61.37
27.0	7.53	41.0	86.86
27.0	7.53	41.0	64.18
28.0	16.06	41.0	64.18
28.0	10.39	41.0	69.85
30.0	10.39	41.5	67.04
32.0	24.54	41.5	64.18
33.0	27.35	41.5	55.70
33.0	24.54	41.5	64.18
33.0	10.39	42.0	67.04
34.0	30.21	42.0	67.04
34.0	35.88	42.0	61.37
34.0	33.02	42.0	64.18
34.0	35.88	42.5	64.18
35.0	33.02	42.5	64.18
35.0	16.06	42.5	67.04
36.0	55.70	42.5	64.18
38.0	55.70	42.5	64.18
38.0	58.51	42.5	64.18
38.5	55.70	42.5	58.51
38.5	52.84	42.5	75.52
39.0	69.85	43.0	64.18
39.0	75.52	43.0	84.00
39.0	52.84	43.0	69.85
39.0	58.51	43.0	67.04
39.0	58.51	43.0	64.18
39.0	69.85	43.5	78.33
39.5	58.51	44.0	81.19
40.0	64.18	44.0	75.52
40.0	67.04	44.5	72.67
40.0	58.51	44.5	81.19
40.0	67.04	45.0	84.00
40.5	64.18	45.0	72.67
41.0	67.04	45.0	81.19
41.0	58.51	45.0	72.67
41.0	64.18		

assumptions, are fewer than they are when a zooarchaeologist seeks to determine weights of useable meat from biomass estimates. Of course, the analytical goal and thus the target variable(s) of the paleontologist are different than those of a zooarchaeologist. The paleontological target variable often concerns something other than taxonomic abundances, such as the average body mass of adult members of a taxon to gain insight to physiology or metabolism or the like. One begins with one or more measurements of a bone (or several bones) and then estimates body mass based on bone size. An example will make the analytical protocol clear.

Based on a sample of seventy-one white-tailed deer (*Odocoileus virginianus*), Emerson (1978) argued that the length of the astragalus would provide fairly accurate estimates of the live weight of individual deer. He weighed field-dressed carcasses, converted those weights to live weights using a “commonly accepted regression equation” developed by wildlife biologists (Emerson 1978:36), measured the length of an astragalus from each carcass, and then a regressed astragalus length against (estimated) live weight. His data are presented in Table 3.12, and a bivariate scatterplot of those data and the statistical results of his analysis are given in Figure 3.6. I converted Emerson’s weights, which were given in pounds, to kilograms, and I reversed the axes in Figure 3.6 from what Emerson (1978:42; 1983:66) originally illustrated because in an archaeological setting the length of the astragalus would be the independent variable and carcass weight would be the dependent variable. Emerson suggested that the regression equation he derived from the data could be used to predict the live weight of individual deer represented in an archaeological collection of astragali. The equation I derived from those data ( $Y = -104.96 + 4.11 X$ ; where  $Y$  is the live weight or biomass in kilograms, and  $X$  is the length of the astragalus in millimeters) could also be used in this fashion, with certain caveats.

Later workers were explicit about the fact that using linear allometry to measure biomass provided at best ordinal scale data, even when sex and age could be controlled (Purdue 1983, 1987). Purdue (1987:10) was particularly insightful when he observed that an estimate of biomass provided by linear allometry “should be viewed as an index that smooths multiple compounding factors and is useful only in an ordinal sense.” Purdue (1987) derived several measurements of biomass from multiple skeletal elements and calculated an average biomass per individual animal, further smoothing his results. Some researchers argue that proximal limb elements of mammals, such as the humerus and femur, should be used rather than distal limb elements such as the radius-ulna, tibia, and metapodials because proximal limb elements produce higher coefficients of determination describing the relationship between a linear dimension of a bone and body weight (McMahon 1975; Noodle

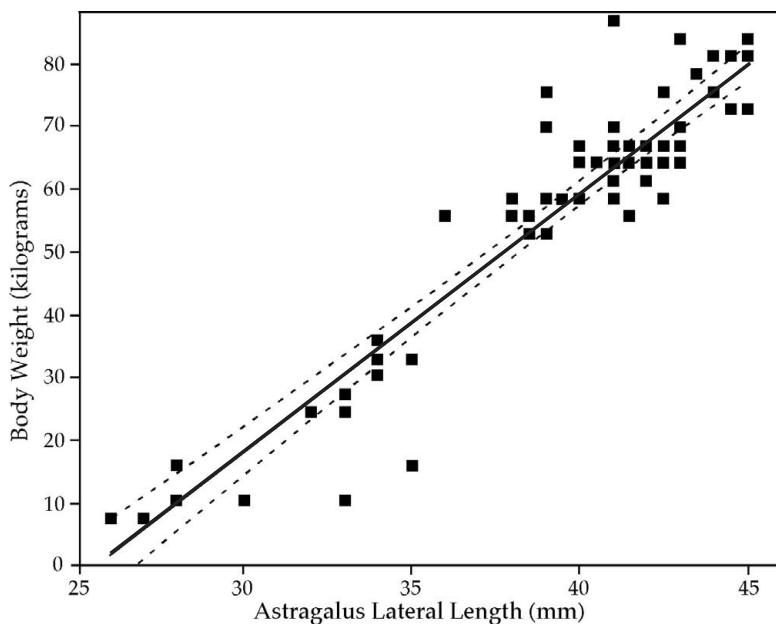


FIGURE 3.6. Relationship between lateral length (millimeters) of white-tailed deer astragali and body weight (kilograms). 95 percent confidence interval indicated by dashed lines. Some points represent multiple specimens. Data from Table 3.12.

1973; Reitz and Honerkamp 1983). This may be so, but we need only worry about it if we desire ratio scale results, and such results are impossible to attain with any biomass (or usable meat) measurement technique.

Before we leave the subject of bone size, a final point needs to be made. As with the skeleton's weight being closely related to the body's biomass, many (but not all) linear dimensions of skeletal elements are also correlated with body size (Orchard 2005, and references therein). Paleozoologists can exploit that relationship in one of two ways. One is to match bones (are a left and a right specimen members of a bilateral pair or not) based on the principle that unique skeletal elements from the same individual will predict essentially the same body size (Nichol and Creak 1979); those prehistoric specimens that suggest they are from the same size animal could be matched or inferred to be from the same animal. This exploits bone size to control for skeletal element interdependence and allows anatomical refitting. In a similar sort of analysis, different skeletal elements that represent a unique body size could each be inferred to represent a unique individual, thereby increasing the total number of individuals to a value likely to be greater than a standard MNI

(Orchard 2005). However, one difficulty here is determining when a difference in body size (indicated by difference in bone size) also represents a different individual. Furthermore, aggregation will influence tallies of individuals because it is likely that bilateral pairs and nonmatching specimens will only be sought within each aggregate. Because of these difficulties, results derived using these procedures will be at best ordinal scale measures of taxonomic abundances.

Paleozoologists can also use the relationship of bone size to body size to monitor clines (character gradients) over time and space. And they can do so without converting bone size to an estimate of body size or body mass (e.g., Butler and Delacorte 2004; Lyman 2004b; Lyman and O'Brien 2005; Purdue 1980, 1986). Even then, however, it would likely be imprudent to suggest that a shift represented, say, a 5 percent decrease in size because that decrease is best considered ordinal scale. If the shift were based on averages, much variation would be smoothed. Use of skeletal element size would not provide quantitative information like measures of taxonomic abundances or of biomass. But, as with many quantitative measures discussed thus far, the research question (or problem) one asks dictates a particular target variable, and whether or not we can design a measured variable that correlates strongly with that target variable is the challenge. Biomass, usable meat, and consumed meat are similar sorts of measures, sometimes derived from a collection of faunal remains with very similar techniques. Given that they are based on the identified assemblage (Figure 2.1), how do they relate to a target variable?

More than 35 years ago, paleozoologist John Guilday (1970) determined the MNI represented by a sample of remains recovered from the site of an historic fort. Historical documents indicated that the fort had been occupied continuously for about 2,364 days by anywhere from 8 to 4,000 men. The faunal remains represented approximately 1,815 kilograms of meat. Guilday (1970) noted that at a standard field ration of about a half kilogram of meat per man per day, the site could have been occupied by 4,000 men for one day, or by only two men the entire time it was in fact occupied. He concluded that calculations of meat weight were “patently ridiculous.” There is an important lesson here. Given that archaeologists usually excavate only part of a site and thus collect but a sample of the faunal remains in the site deposits, why would anyone want to try to calculate meat amounts? A partial answer seems to be that many believe meat amounts provide ordinal scale estimates of which taxa provided the most food and which provided the least. Whether those amounts are in fact ordinal scale is usually assumed rather than tested. To perform a test, examine the magnitude of difference between taxon specific amounts (e.g., Figures 3.4–3.5).

## UBIQUITY

What is termed a “ubiquity index” has long been used in paleoethnobotany (Popper 1988). Ubiquity concerns the frequency of (depositional) contexts in which a taxon occurs and thus it is measured in several ways. In the simplest way, the absolute frequency of distinct archaeological features in a particular archaeological context that contains a taxon is tallied, and the total is a measure of the ubiquity of that taxon in the archaeological context under consideration (e.g., Purdue et al. 1989; Stahl 2000). A variant of this procedure is to determine the percentage of assemblages (whether of sites, components, strata, or features) that contain remains of a taxon (e.g., Lubinski 2000). The other major way to measure ubiquity is to construct a bivariate scatterplot with the number of archaeological contexts containing remains of a taxon on one axis and the NISP of the taxon on the other axis (Styles 1981:43). Styles (1981:44) cautions that the bivariate scatterplot “is not a direct measure of taxon importance” because it in part measures human behaviors as well as natural taphonomic processes. The bivariate scatterplot also allows visual assessment of intrataxonomic variation in ubiquity and simultaneous assessment of the influence of sample size on ubiquity. If some taxa are more ubiquitous than others with similar sample sizes, it is reasonable to conclude that some taphonomic agent or process accumulated and deposited, or dispersed remains of one taxon differently than another. Identification of the agent or process is a taphonomic issue (Lyman 1994c).

The ubiquity index can be calculated in other ways. These tend to be derivative of the first technique. Consider the taxonomic abundance (NISP) data from the collection of eighty-four owl pellets in Table 2.8. Note that *Sylvilagus* is represented in two pellets, *Reithrodontomys* is represented in four pellets, *Sorex* in seven pellets, *Thomomys* in eleven pellets, *Microtus* in forty-nine pellets, and *Peromyscus* in sixty pellets. If ubiquity of a taxon is measured as the total number of pellets in which a taxon occurs, then the relationship between NISP and ubiquity of a taxon is very strong and significant (Figure 3.7;  $r = 0.997$ ,  $p < 0.0001$ ). That shouldn’t be surprising. On the one hand, a taxon can occur in no more pellets (or any other context) than its NISP, so, for example, *Sylvilagus* can occur in only five pellets because it has an NISP of five, and *Reithrodontomys* can occur in no more than nineteen pellets because it has an NISP of nineteen. In other words, the number of contexts (Ncontexts) in which a taxon occurs or ubiquity  $\leq$  NISP. A taxon with NISP  $>$  Ncontexts, on the other hand, can occur in as many as Ncontexts; it is not limited in its ubiquity.

The most obvious other way to measure ubiquity is to tally up the number of sites in a region that contain remains of a taxon (e.g., Butler and Campbell 2004). Or, tally up the number of collections or temporally distinct assemblages within a single

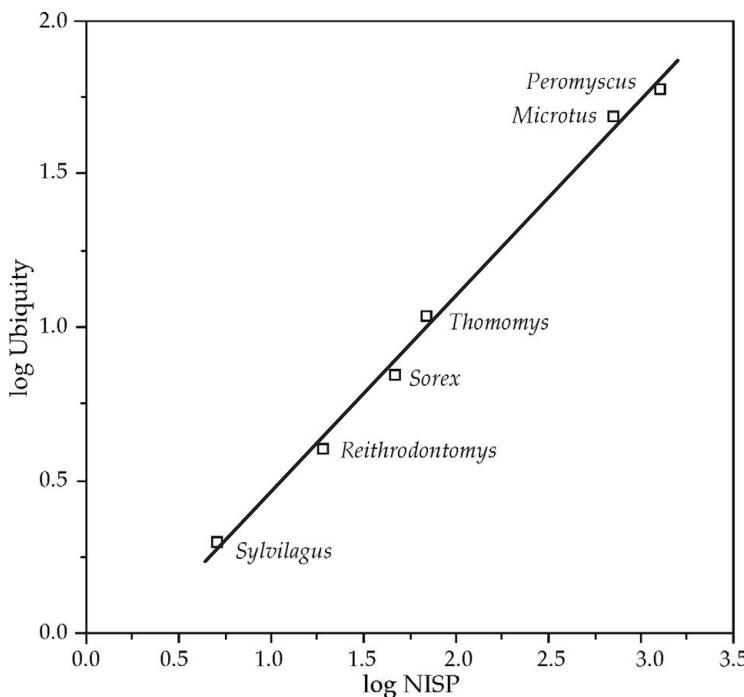


FIGURE 3.7. Relationship between NISP and ubiquity (number of pellets in which a taxon occurs) of six genera in a collection of eighty-four owl pellets. See Table 2.8.

site in which a taxon occurs. However, any such tally, regardless of the spatial or temporal scale at which ubiquity is measured, may well reflect sample size. Consider the data for eighteen sites in eastern Washington State. Fourteen of these sites were used in analyses discussed in Chapter 2; four additional sites are included here to increase the number of assemblages examined. The taxa (all mammals) represented are unimportant to this exercise. What is important is that all eighteen collections are from a single stretch of river about 45 km long. There are minor habitat differences among the sites, and some age variation but all date to the last 7,000 years. All sites have the same probability of producing the same taxa. Thus, all taxa *should* be equally ubiquitous across these sites if every site was sampled in a manner equivalent to every other site (all else being equal, such as accumulation, preservation, and recovery). The only difference in sampling across the sites was the volume of sediment excavated, and thus, not surprisingly, the total NISP per site varies. The last leads to the prediction that taxa with many NISP will tend to be more ubiquitous than taxa with few NISP. Do the data meet this prediction?

Both NISP and ubiquity data for the twenty-eight taxa represented are presented in Table 3.13. NISP values per taxon have been summed across all 18 sites, and range

Table 3.13. *Ubiquity and sample size of twenty-eight mammalian taxa in eighteen sites.  $\sum$ NISP is the total NISP per taxon for all sites summed. Ubiquity is the number of sites in which remains of a taxon occur*

Taxon	$\sum$ NISP	Ubiquity	Taxon	$\sum$ NISP	Ubiquity
1	37	8	15	106	15
2	77	3	16	6	1
3	176	15	17	9	2
4	35	9	18	14	2
5	793	17	19	10	2
6	95	13	20	9	5
7	14	6	21	63	7
8	23	4	22	10	3
9	1,022	18	23	7	3
10	140	16	24	273	11
11	75	10	25	24	2
12	29	8	26	107	11
13	2,706	18	27	10,062	17
14	4	3	28	1,953	14

from a low of 4 to a high of 10,062 per taxon. Ubiquity ranges from one to eighteen. The relationship between log NISP per taxon and log ubiquity per taxon is statistically significant ( $r=0.802, p<0.0001$ ). The relationship between the two is described by the equation  $Y=0.192X^{0.34}$  and is shown in Figure 3.8. Across these eighteen assemblages ubiquity is strongly related to sample size measured as NISP. This means that if more of each site with low NISP values had been excavated, it is likely that they would have eventually produced not only more NISP but also more of the taxa documented in nearby sites that they currently lack. Thus, to say that in the area and time range sampled by these eighteen collections, some taxa were relatively ubiquitous (for whatever reason) whereas other taxa were not very widespread and had low ubiquity might be correct, but it might also be incorrect in the sense that ubiquity is a function of sample size measured as NISP. Taxa that do not occur in many collections may be nonubiquitous because an insufficient amount of excavation has been done at some sites and thus few of the remains of these less ubiquitous taxa were recovered.

But one might argue that taxa represented by a total NISP  $< 18$  could not possibly have a ubiquity of eighteen. The ubiquity of a taxon can be no greater than that taxon's  $\sum$ NISP across all recovery contexts, and that mechanical truism may be influencing statistical results. If we omit all taxa in Table 3.13 with NISP  $< 18$ , the correlation between log NISP and log Ubiquity for the remaining nineteen taxa is a bit

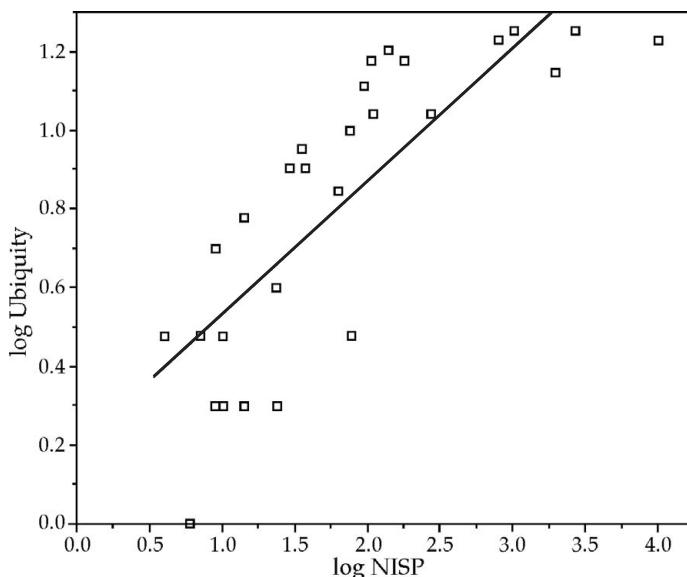


FIGURE 3.8. Relationship between NISP and ubiquity of twenty-eight taxa in eighteen sites. Data from Table 3.13.

weaker than when all twenty-eight taxa are included, but that correlation coefficient is still significant ( $r = 0.67, p = 0.002$ ). Thus it would be unwise to argue that more ubiquitous taxa were “more important” than those that are less ubiquitous. In this set of eighteen collections we cannot discount the possibility that ubiquitous taxa are ubiquitous because excavations produced many specimens of each; nonubiquitous taxa are not ubiquitous because relatively few specimens of each were recovered. Those rarely represented taxa may not be very ubiquitous because they are rare on the landscape, or they were rarely accumulated, or their remains were rarely preserved, or rarely recovered. How to determine which of these possibilities holds in any given case demands taphonomic analysis. Such analysis is oftentimes difficult when remains are few in number, which is of course the case when taxa are not ubiquitous.

In the case described in the preceding paragraph ubiquity was measured across multiple sites, but ubiquity can also be measured across analytical units or strata or features within a single site. In these cases, too, any measure of ubiquity is prone to be larger with larger sample sizes (greater NISP values). This can be shown using two of the multicomponent collections in the eighteen-site sample. Data for these two collections are given in Table 3.14. In both sites the ubiquity of taxa is strongly related to sample size. Site 45DO189 has seven analytical units and fifteen total mammalian taxa (Lyman 1988). Ubiquity is significantly correlated with NISP ( $r = 0.662, p = 0.007$ ). Site 45OK2 has four analytical units and eighteen total mammalian taxa (Livingston

Table 3.14. *Ubiquity and sample size of mammalian taxa across analytical units in two sites.*

45OK2: taxon	NISP	Ubiquity ( $N_{\max} = 4$ )	45DO189: taxon	NISP	Ubiquity ( $N_{\max} = 7$ )
1	6	2	1	1	1
2	7	3	2	6	4
3	6	2	3	86	6
4	27	3	4	14	5
5	9	2	5	1	1
6	2	1	6	13	3
7	1	1	7	4	3
8	110	4	8	10	2
9	14	4	9	6	4
10	272	4	10	1	1
11	7	3	11	1	1
12	6	1	12	2	2
13	2	2	13	251	7
14	1	1	14	5	2
15	16	4	15	14	6
16	7	4			
17	2,021	4			
18	51	3			

1984). Here, ubiquity is also significantly correlated with NISP ( $r = 0.709, p = 0.001$ ). The relationship of sample size and ubiquity across analytical units at 45DO189 is illustrated in Figure 3.9; that relationship as it is manifest at 45OK2 is shown in Figure 3.10. Again, some taxa may indeed be more ubiquitous than others in the sense of being found associated with more spatiotemporal analytical units. But it is difficult to make this argument on empirical grounds because the available data suggest that had more of each analytical unit with low NISP values been excavated, NISP would have been larger, more taxa would have been found in those units, and the ubiquity of rarely represented taxa (those with low NISP values) would have increased.

Dean (2005a:416–417) cited an ethnobotany text as indicating that ubiquity measures minimize influences of “random variations [in] NISP counts” when seeking to measure taxonomic abundances, and those measures also allow comparisons of assemblages from different habitats, assemblages collected using different recovery procedures, and assemblages subject to varied preservation. To be sure, sampling different habitats will influence NISP, as will differential preservation and recovery. But in so far as ubiquity is a function of NISP – and the examples given above suggest

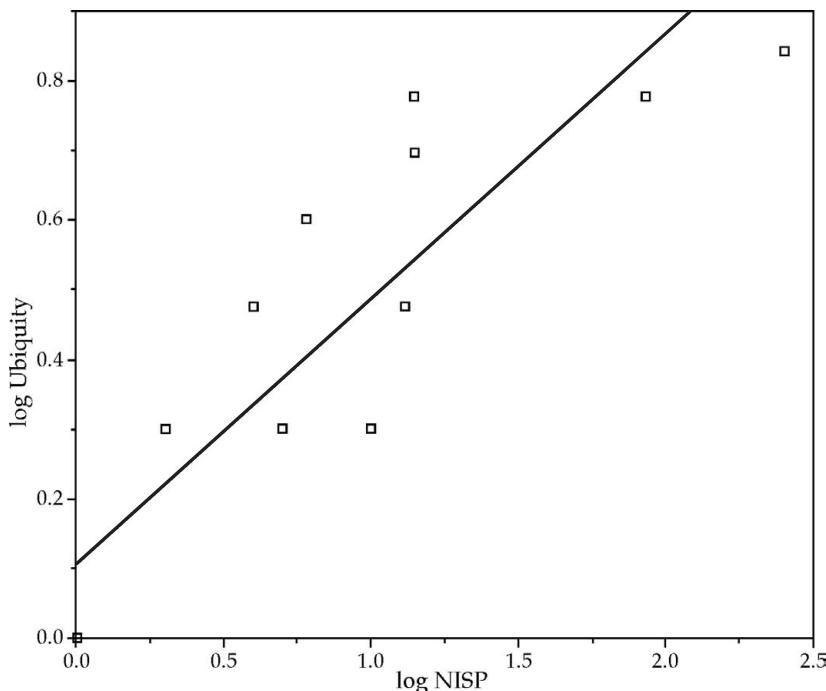


FIGURE 3.9. Relationship between NISP and ubiquity of fifteen taxa in seven analytical units in site 45DO189. Data from Table 3.14.

that ubiquity will often be a function of NISP – differential sampling intensity, differential preservation, and differential recovery will also influence ubiquity because they influence NISP (see also Kadane 1988).

The preceding is not to say that ubiquity measures are valueless. These could be quite useful if the influences of sample size could be controlled. If, say, the NISP values of several taxa are very similar (say within 5 percent of each other), and one taxon has a high ubiquity value and another has a low value, then it would be reasonable to suspect that some mechanism or agent of accumulation had dispersed widely the remains of the former taxon and perhaps that same mechanism or agent (or another one) had concentrated (or failed to disperse) the remains of the latter taxon. Given this possibility, it is difficult to understand why ubiquity has not been measured more frequently.

#### MATCHING AND PAIRING

Recall the definitions of MNI provided by Stock, Howard, Adams, and the ornithologists interested in raptor diets (and see Table 2.4). White (1953a:397) defined MNI

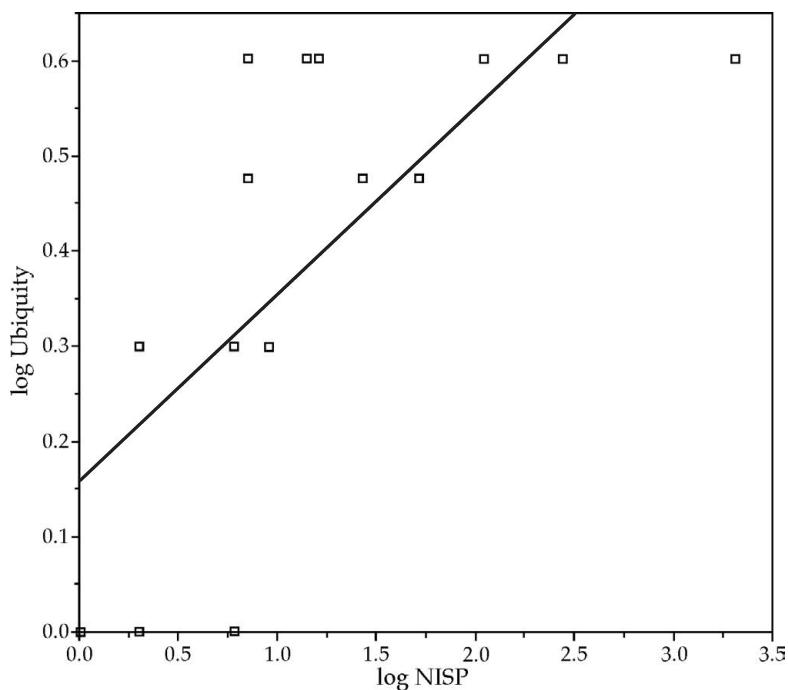


FIGURE 3.10. Relationship between NISP and ubiquity of eighteen taxa in four analytical units in site 45OK2. Data from Table 3.14.

this way: The “number of individuals represented by the excavation sample [is determined as follows.] Separate the most abundant element of the species found . . . into right and left components and use the greater number as the unit of calculation” (see also White 1953b:61). White went on to note that this procedure would produce a “slight error on the conservative side because, without the expenditure of a great deal of time with small returns, we cannot be sure all of the lefts match all of the rights” (1953a:397).

What White was getting at with the idea of “matching” is this. Recall that when faced with three left and two right scapulae of a taxon, the number of individuals is a *minimum* because at least three individuals had to contribute these bones. Perhaps as many as five individuals contributed the scapulae, but unless we can show that the two rights do not “match” or bilaterally pair with any of the three left scapulae, then we must conclude that the two rights are from two of the three individuals represented by the lefts. How do we make a determination of whether any of the lefts match any of the rights? Usually *bilaterally paired bones* are compared, when these are elements with left and right members, such as scapulae, humeri, tibiae, and so on. Skulls are not paired bones, but mandibles are; vertebrae are not, but ribs

are, although ribs are seldom matched because they tend to not be taxonomically diagnostic beyond the family level (if that).

To determine if two potentially paired bones indeed “match” or are bilaterally paired, they must be compared under the assumption of bilateral symmetry. The two members of the possible pair (obviously from the same species) are compared in terms of their size, their ontogenetic (growth) development, the sex of the individual represented, and the like (Chaplin 1971:70). Because vertebrates, such as mammals and birds, are more-or-less bilaterally symmetrical, the left bone should closely match the right bone in terms of all characteristics if they are from the same individual. White thought that matching would gain little, but provided little data to this effect. It is legitimate to ask, then: What does identifying bilaterally paired bones gain in terms of measures of taxonomic abundance? In fact, can bilateral pairs be accurately identified? To answer these questions, the basics of the matching procedure and how it influences measures of taxonomic abundance are reviewed first. Then a study of how accurate identifications of bilateral mates might be is presented.

### *More Pairs Means Fewer Individuals*

In his seminal example, Chaplin (1971:71–75) described a fictional example of how identifying pairs would provide a different number of individuals than a Whitean MNI. He did not reference White, but Chaplin (1971:70) did note that a Whitean MNI – one defined by the most common skeletal element or part – would constitute a “not necessarily very satisfactory estimate of the true minimum.” If one had, say, eleven distal left humeri and five proximal right humeri, and none of these specimens overlapped anatomically, then the Whitean MNI would be eleven but Chaplin’s point was that there might actually be twelve, thirteen, fourteen, fifteen, or even sixteen individuals represented if some of the proximal right specimens were not from the same animals as the left distal specimens. Thus he focused on identifying those specimens without matches; matches would not increase the MNI, but bones without matches would increase that number because they would be added to the number of pairs or matches. Thus, more pairs mean fewer individuals and fewer pairs mean more individuals. Left specimens without a mate and right specimens without a mate are added to the number of pairs as independent representatives of individual organisms. The more pairs of left and right elements, the fewer independent left and right elements added to the total. White’s method of determining MNI assumes all lefts pair up with a right element, and all rights pair up with a left element. Chaplin was unwilling to make this assumption.

Chaplin's (1971) example of how matching influences measures of taxonomic abundance involved domestic sheep (*Ovis aries*) tibiae. He used fictional data, but those data illustrate the issues involved nicely. The NISP of left specimens was thirteen; some specimens were anatomically complete left tibiae, some represented various portions of left tibiae. The (minimum) number of skeletal elements represented (based on age, sex, size) was ten, which also represented the MNI for left tibiae. The NISP of right tibia specimens in Chaplin's example was seventeen; these represented fifteen elements and fifteen MNI. Based on age (epiphyseal fusion) and size (see below), Chaplin identified bilateral pairs of left and right tibiae. He derived a formula for calculating what he called the "grand minimum total" or GMT of individuals. This formula is:

$$GMT = (C^t/2) + D^t,$$

where  $C^t$  is the total number of skeletal elements (not specimens) making up bilaterally matched pairs and  $D^t$  is the total number of elements (not specimens) without bilateral mates. In his example, Chaplin identified eight pairs comprising sixteen elements (eight lefts and eight rights), and had two unmatched lefts and seven unmatched rights. Thus,  $C^t = 16$ ,  $D^t = 9$ , and  $GMT = (16/2) + 9 = 8 + 9 = 17$  individuals. Thus, in this example, the Whitean "most common element" MNI was fifteen (based on right tibiae), but matching bilateral elements indicated that because some elements lacked their bilateral mate and represented independent organisms, a more accurate (but still a minimum) tally of individuals was seventeen.

A Whitean "most common element" MNI assumes that each left humerus, say, has a bilateral mate among the right humeri, so only the lefts, or only the rights, but not both lefts and rights contribute to the MNI tally (see Table 2.5). Chaplin's GMT can produce larger tallies of individuals because both left humeri and right humeri can contribute to the tally. Matching establishes that some left humeri are independent (come from a different individual organism) of all right humeri, and some right humeri are independent of all left humeri. Therefore, two "most common elements" – left and right humeri – are identified rather than either left or right humeri only. Calculating GMT for a collection will give a larger number of individuals the fewer the identified pairs. In Chaplin's example, if we reduce the number of pairs to 7 ( $C^t = 14$ ), that increases  $D^t$  to 11, so  $GMT = (14/2) + 11 = 7 + 11 = 18$ , simply because we have added one more independent element (another left, or right, tibia in this example) to the tally. This might make GMT attractive, but do not be fooled by larger numbers. To be sure, GMT can produce larger numbers of individuals than the Whitean "most common element" MNI. But does that gain us anything with respect to measuring taxonomic abundances?

First, the absence of pairs means that MNI = NISP. GMT provides measures of taxonomic abundance between NISP and Whitean MNI values. In Chaplin's example, NISP = 30, GMT = 17, and (Whitean) MNI = 15. Given this, and arguments about the statistical relationship between MNI and NISP, GMT will be a function of NISP. Taxonomic abundances based on GMT will also at best be ordinal scale. Think of GMT as simply a different way to aggregate faunal remains to derive MNI values. GMT values will depend heavily on aggregation, just as MNI does. This is so because it is unlikely that one would seek to identify bilateral pairs (or a lack thereof) among assemblages of tibiae from different strata deposited at different times (unless one was interested in postdepositional disturbance processes that resulted in inter-strata movement of specimens). Rather one would choose to seek a bilateral pair among skeletal elements that potentially derive from the same animal. Thus, how one defines aggregates plagues GMT just as it does MNI.

Another problem with determination of GMT concerns identifying bilateral pairs. Because this problem afflicts all measures of taxonomic abundance that use bilateral pairs in the calculation, it is addressed later in this chapter. The bottom line to Chaplin's GMT should be clear. Although GMT tries to make quantitative use of independent left and right skeletal elements, it is plagued by many of the same difficulties as Whitean MNI values. That it produces data with no more quantitative resolution or validity than NISP with respect to measures of taxonomic abundances should cause one to pause before calculating GMT, if it is calculated at all.

### *The Lincoln–Petersen Index*

Chaplin's (1971) GMT is not the only quantitative procedure meant to measure taxonomic abundances that uses bilateral pair data (e.g., Krantz 1968; Lie 1980, 1983; Wild and Nichol 1983a, 1983b; Winder 1991). A couple of these other techniques have undergone critical evaluation (Allen and Guy 1984; Bokonyi 1970; Casteel 1977; During 1986; Fieller and Turner 1982; Gautier 1984; Horton 1984), and are seldom used these days. They are not considered further. But there is one method that requires comment because it has several supporters, it has been suggested at least twice by independent workers (Allen and Guy 1984; Fieller and Turner 1982), and it might well be suggested yet again given that versions of it are used by wildlife biologists today (e.g., Amstrup et al. 2006; Hopkins and Kennedy 2004; Slade and Blair 2000). More importantly, biological anthropologists have recently suggested it is useful for estimating the number of individual humans in commingled assemblages of remains (Adams and Konigsberg 2004).

The most frequently advocated method of using identified bilateral pairs in the service of measuring taxonomic abundances is analogous to (and in fact derived from) a procedure used by wildlife biologists to estimate taxonomic abundances. In wildlife biology, it is known as “capture–recapture analysis,” where a sample of animals is captured, tallied, and each captured individual ( $n_1$ ) is marked ( $m_1$ ) and released (Nichols [1992] provides a good introduction). Then a second sample of animals ( $n_2$ ) is captured. The number of previously marked ( $m_1$ ) individuals that are recaptured ( $m_2$ ) and the number of unmarked (new) individuals ( $n_2$ ) in the second sample are tallied. These values –  $n_1$ ,  $n_2$ ,  $m_1$ ,  $m_2$  – are used to estimate the size of the population from which the two samples were drawn. The quantitative measure is usually referred to as the Petersen index, and less often as the Lincoln index (Fieller and Turner 1982; Turner 1980, 1983); it is here referred to as the Lincoln–Petersen index. Several paleozoologists find this index to be superior to Whitean MNI values and also superior to Chaplin’s GMT values (e.g., Allen and Guy 1984; Fieller and Turner 1982; Turner 1980, 1983; Turner and Fieller 1985; Winder 1991).

In the Lincoln–Petersen index,  $n_1 = m_1$ . After release of the marked individuals comprising the first sample, the proportion ( $P$ ) of marked individuals in the population can be symbolized as  $n_1/Y = P$  or  $m_1/Y = P$ , where  $Y$  is the population size. What we of course do not know but seek to estimate is  $Y$ . So, we convert  $m_1/Y = P$  first to  $m_1 = PY$  (multiply both sides of the equals sign by  $Y$ ), then convert the latter to  $m_1/P = Y$  (divide both sides of the equals sign by  $P$ ). The Lincoln–Petersen index assumes that the proportion of marked individuals in the second sample effectively estimates the proportion of marked individuals in the population, or  $P = (m_2/n_2)$ . Substituting  $(m_2/n_2)$  for  $P$  into  $Y = m_1/P$ , the result is

$$Y = m_1/(m_2/n_2).$$

Because division by a fraction is equivalent to multiplying by the reciprocal of that fraction,

$$Y = m_1 n_2 / m_2.$$

And because  $m_1 = n_1$ , the preceding equation can be rewritten as

$$Y = n_1 n_2 / m_2.$$

Either of the last two formulae provides an estimate of the size of the population from which the individuals comprising the two samples ( $n_1$  and  $n_2$ ) were drawn. The reasoning is illustrated in Figure 3.11.

Zooarchaeologists who advocate use of the Lincoln–Petersen index for estimating taxonomic abundances observe that bilaterally paired skeletal elements, such as left

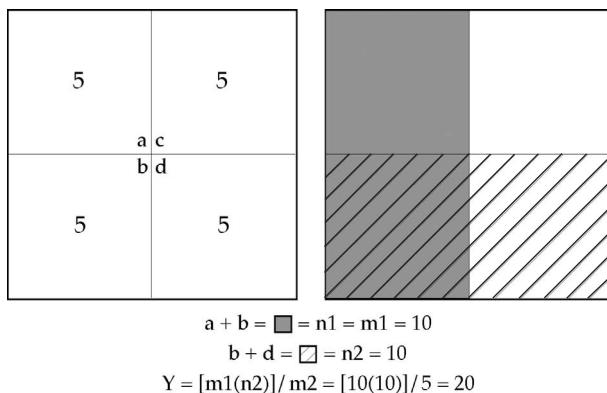


FIGURE 3.11. A model of how the Lincoln–Petersen index is calculated. Square on the left is the population (twenty individuals) divided into four subpopulations (a–d). Square on the right illustrates the captured and marked first sample ( $n_1$  or  $m_1$ ; shaded), and the second sample ( $n_2$ ; cross-hatched) comprising some marked individuals from the first sample (cell  $b = m_2$ ) and individuals captured for the first time (cell  $d$ ).

and right femora, can be used as follows. The lefts can be considered one sample (say,  $n_1$ ), and the rights can be considered the other sample (say,  $n_2$ ). Identifying bilateral pairs of left and right elements is analogous to finding a marked individual in the second sample. If left elements are treated as the first sample,  $L = n_1$ ; if right elements are treated as the second sample,  $R = n_2$ ; and if the number of paired bones is treated as the number of marked recaptures,  $m_2$  = number of bilateral pairs, then we can substitute these symbols into the Lincoln–Petersen index to estimate the number of individuals in the population from which the bones came. The formula can be written as

$$LP_{\text{ind}} = L R / m_2,$$

where  $LP_{\text{ind}}$  is the estimated number of individuals given by the Lincoln–Petersen index,  $L$  is the number of left elements,  $R$  is the number of right elements, and  $m_2$  is the number of bilateral pairs. In wildlife biology, by convention, if no pairs are found,  $LP_{\text{ind}}$  is estimated as  $LP_{\text{ind}} = (n_1 + 1)(n_2 + 1)$ . And by convention,  $LP_{\text{ind}}$  is rounded up to the nearest whole number.

$LP_{\text{ind}}$  provides an estimate of taxonomic abundance that is greater than MNI and GMT. The calculated value of  $LP_{\text{ind}}$  for a taxon in any given assemblage depends on the NISP for that taxon and also on the number of bilateral pairs of skeletal elements of that taxon regardless of NISP. In the fictional data in Table 3.15, for the sake of simplicity, each specimen is an anatomically complete skeletal element. As NISP ( $= L + R$ ) per row increases in Table 3.15, if the number of pairs ( $m_2$ ) remains

Table 3.15. Fictional data illustrating influences of NISP and the number of pairs on the Lincoln–Petersen index ( $LP_{\text{ind}}$ )

$\Sigma \text{NISP}$	Left ( $n_1$ )	Right ( $n_2$ )	Pairs ( $m_2$ )	$LP_{\text{ind}}$
20	10	10	2	50
20	10	10	3	34
20	10	10	4	25
20	10	10	5	20
20	10	10	6	17
20	10	10	7	15
20	10	10	8	13
40	20	20	3	134
60	30	30	3	300
80	40	40	3	533
100	50	50	3	834

stable, then the associated  $LP_{\text{ind}}$  increases. As NISP increases from 20, to 40, to 60, to 80, to 100, if  $m_2 = 3$ , then  $LP_{\text{ind}}$  increases from 34, to, 134, to 300, to 533, to 834, respectively. And, if the NISP per row remains stable but  $m_2$  increases, then  $LP_{\text{ind}}$  decreases. Consider those rows in which  $\Sigma \text{NISP} = 20$ . As  $m_2$  increases from 2 to 8 by increments of 1,  $LP_{\text{ind}}$  decreases from 50, to 34, to 25, to 20, to 17, to 15, to 13. Increase in NISP causes the estimated number of individuals to increase (all else [ $m_2$ ] equal); increase in the number of pairs causes the estimated number of individuals to decrease (all else [NISP] equal) until all bones are paired in which case  $m_2 = LP_{\text{ind}}$ .

Fieller and Turner (1982) identify several properties of  $LP_{\text{ind}}$  that they argue are beneficial to estimates of taxonomic abundance. One is that confidence intervals can be calculated for any value; as they put it, “it is possible to specify a range of plausible values for the total population size that are ‘reasonably’ consistent with the observed proportion of tagged animals” or paired bones (Fieller and Turner 1982:53). Calculation of both  $LP_{\text{ind}}$  and its associated confidence interval for a taxon rests on several assumptions that, if violated, can cause the estimate to be far off the mark. For example, the capture–recapture-based Lincoln–Petersen index assumes that in the time between the release of marked individuals ( $n_1 = m_1$ ) and the collection of the second sample ( $n_2$ ), the marked individuals will be *randomly mixed* into the population. That is, it is assumed that the probability of drawing a marked individual in  $n_2$  will not be influenced by the fact that that marked individual was also a part of the first sample or  $n_1$ . Despite more than 25 years of detailed ethnoarchaeological studies of faunal remains, only a small bit of that research has focused on the spatial

distribution of those remains (e.g., Bartram et al. 1991), and only a small fraction of that research has focused on matching (e.g., Waguespack 2002). We do not yet know that the requisite assumption is valid, but the limited available data suggest that it will not always be. Faunal remains are seldom randomly distributed.

Some suggest the benefit of  $LP_{ind}$  is that it “accounts for random data loss” (Adams and Konigsberg 2004:140). In taphonomic terms, whether or not a left humerus, say, is accumulated or preserved should not influence the accumulation and preservation of its bilateral (right) mate. Fieller and Turner (1982) argue that calculation of multiple  $LP_{ind}$  values and their respective confidence intervals based on different skeletal elements – humeri, radii, femora, and so on – should produce similar estimates of a taxon’s abundance if the assumptions requisite to its calculation are not violated. The only thing these similar values indicate is that the included paired skeletal elements have spatial distributions and preservation potentials sufficiently similar that each provides a similar  $LP_{ind}$  value. Whether left and right femora are randomly distributed relative to each other and also relative to both left and right humeri is a separate question. This distinction is confirmed by Fieller and Turner’s (1982:55) suggestion that dissimilarity in  $LP_{ind}$  values for different elements reveals something about “differential deposition [that is] of considerable interest.” Do not confuse deposition in site sediments with deposition in the area excavated. We do not need to calculate  $LP_{ind}$  values to determine that the parts of skeletons are differentially represented, whether due to varied deposition or preservation. Does  $LP_{ind}$  provide something not provided by other measures of taxonomic abundances?

Fieller and Turner (1982:54) argue that  $LP_{ind}$  is the only quantitative unit that provides an “estimation of the original killed population size.” According to Fieller and Turner (1982:51), MNI (of either the Whitean most common element type, or Chaplin’s GMT type) provides a “simple count” of the number of individuals necessary to account for the bones on the lab table. They argue that  $LP_{ind}$ , in contrast, provides an estimate of the original killed population because it takes into account “missing material” and thus it is a “conventional statistical estimate” (Fieller and Turner 1982:51). Ringrose (1993:129) agrees that the Lincoln–Peterson index provides estimates of taxonomic abundances within the “death assemblage.” And Adams and Konigsberg (2004:138) seem to agree but say the method “estimates the *original* number of individuals represented by the osteological assemblage” (emphasis in original). But consider whether the population of twenty individuals shown in Figure 3.11 comprises the biocoenose, the thanatocoenose, or the taphocoenose. Recall Figure 2.1, where the differences between a biocoenose (living population on the landscape), a thanatocoenose (population of dead individuals, or killed population), a taphocoenose (what is accumulated, deposited, and preserved in a site), and an identified

assemblage (what was recovered, identified, and reported) are shown graphically. Fieller and Turner (1982) argue that the thanatocoenose from which the identified assemblage is derived is the target variable. They suggest that if MNI or GMT is used the measured variable constitutes the identified assemblage and if  $LP_{ind}$  is used then the estimated variable is the thanatocoenose.

In response to Fieller and Turner's (1982; see also Turner 1983) arguments regarding measured and target variables, and their attendant advocacy of the Lincoln–Petersen index, Grayson (1984:88) pointed out that “an unmatched bone whose partner has simply not been collected has a very different meaning from an unmatched bone whose partner has disappeared.” Winder (1991:126) implied the “very different meaning” was insignificant because both the failure to recover a bone and the lack of preservation of a bone merely concerned “classical sampling theory.” But, Turner (1983:319) himself contradicts this when he points out that “if 100 animals were killed, 60 whole carcasses removed from the site and only a random sample of the bones from the remaining 40 animals deposited, then estimates based on the excavated sample can only refer to the 40 individuals which in this case constitute the killed population.” How can this situation be mathematically different from a case in which 100 animals are deposited in a site but bones of only forty animals preserve and are sampled? How can Turner's example, or the immediately preceding one, be mathematically different from a case in which 100 animals are deposited, but bones of only forty are sampled?

There are no mathematical differences between the three possibilities. This is so because if the drawn sample consists of sixty bones (thirty-five lefts, twenty-five rights) drawn randomly from forty (out of 100) animals with skeletons comprising only one bilateral pair, and there are twenty pairs, the  $LP_{ind} = 44$  regardless of anything else. Certainly those forty-four individuals originated in the biocoenose, passed through the thanatocoenose filter, then through the taphocoenose filter, and finally through the filters of recovery, analysis, identification, and pair matching. Which of those “populations” do those forty-four individuals estimate the size of? That is the key question. Perhaps we do not need an answer for one simple reason. As Fieller and Turner (1982:57) correctly emphasize, “fundamental to the [Lincoln–Petersen] method is the ability to determine the precise number of matches in the assemblage of skeletal parts considered. Omission of true matches inflates the estimate [and] inclusion of false matches has the opposite effect” (see Table 3.15). This, then, comprises another requisite assumption – that analysts can accurately identify bilateral pairs. It is difficult to determine if this assumption is warranted because few published tests of it exist. It is necessary, then, to consider the validity with which bilateral pairs might be identified. If they cannot be consistently validly identified,

then any use of the Lincoln–Petersen index (or any other use of what are thought to be bilaterally paired skeletal elements) is likely invalid.

### *Identifying Bilateral Pairs*

Krantz (1968:287) noted early on that “in experienced hands there should be little doubt as to which mandibles [or any other bilaterally paired elements] pair off and which do not.” Twenty years later, Todd (1987:180) provided a detailed statement on the procedure for identifying bilateral pairs:

Initial estimates of possible anatomical refits can be based on metric attributes of the elements. In the case of bilateral refits, potential candidates for pairs can be further examined visually. Familiarity with elements from individual carcasses allows for the recognition of attributes that are individually distinctive and bilaterally uniform. The patterns of muscle attachment shape and prominence, synovial fovea shape and depth, and the proportions of components of articular surfaces within paired elements in the [animal] carcass create an individually distinctive “finger print” for element mates. Bilateral mates are usually mirror images of each other in these attributes.

Two problems accompany anatomical matching. First, matching takes a considerable amount of time (Adams and Konigsberg 2004; White 1953a); more time is required as the size of the collection (and thus the number of possible pairs) increases. The other problem is more serious. The analyst can determine the sex of the individual represented by paired bones in a mammal skeleton from a very limited number of skeletal elements (innomates, frontals of some antler-bearing ungulates), unless the taxon under consideration is sexually dimorphic. Exacerbating the difficulty of matching is the fact that the degree to which paired bones (and teeth) display the same ontogenetic stage simultaneously is unclear. Thus for illustrative purposes the focus here is bone size. This criterion is used by virtually all who have identified bilaterally paired bones (e.g., Allen and Guy 1984; Chase and Hagaman 1987; Enloe 2003a, 2003b; Enloe and David 1992; Fieller and Turner 1982; Morlan 1983; Nichol and Creak 1979; Todd 1987; Todd and Frison 1992; Turner 1983; Wild and Nichol 1983b).

Chaplin (1971:74) provides an early example of a matching procedure based on bone size. He indicates that left and right distal tibiae of sheep (*Ovis aries*) can be identified as bilateral pairs from the same individual if they are identical in “maximum width,” or if they are within  $\leq 0.3$  mm of each other in maximum width, but lefts and rights are not pairs if their maximum widths differ by  $\geq 0.4$  mm. How he established the 0.3 vs. 0.4 mm maximum width is not specified, but it introduces the

matching problem. Identifying bilaterally paired bones in paleozoological collections rests on the assumption that left and right skeletal elements from the same organism will be symmetrical. Bones (and teeth) are, however, typically asymmetrical to some degree. Indeed, temporally fluctuating asymmetry has in the past two or three decades become a very important research topic in biology (e.g., Gangestad and Thornhill 1998; Palmer 1986, 1994, 1996; Palmer and Strobeck 1986; Pankakoski 1985). To identify bilateral pairs among a commingled set of left and right elements from multiple individuals of a taxon demands that we answer the question “How symmetrical is symmetrical enough to conclude a left and right humerus (for example) are from the same individual” (Lyman 2006a)? Answering that question involves determining a “tolerance” or the maximum allowable asymmetry between bilaterally paired bones (Nichol and Creak 1979). Determination of the tolerance rests on study of known pairs.

Adams and Konigsberg (2004) performed a “test for the accuracy of pair-matching” by choosing skeletons of fifteen humans recovered from an early historic burial ground associated with an archaeological village. Based on morphological indicators such as “robusticity, muscle markings, epiphyseal shape, bilateral expression of periosteal reaction, and general symmetry” (Adams and Konigsberg 2004:145), they successfully identified all pairs of femora, all pairs of tibiae, and all but one pair of humeri. But they also suggested that the error rate might increase were the sample size to increase because large samples would obscure between-individual variation (Adams and Konigsberg 2004:146). As sample size increases (as the number of individuals increases), the degree of discontinuity between individuals will decrease such that one individual will, in a sense, blend into another individual. This will occur with either morphological traits or metric (size) traits.

The bivariate scatterplot in Figure 3.12 shows two measurements of paired distal humeri from forty-eight museum skeletons of deer. Both white-tailed deer (*O. virginianus*,  $N = 17$  pairs) and mule deer (*O. hemionus*,  $N = 30$  pairs) are included; the forty-eighth pair of distal humeri is from a hybrid of the two species. Both species are represented in many paleozoological collections in western North America (Livingston 1987; Lyman 2006b) and their remains can occasionally be distinguished based on morphological criteria (Jacobson 2003, 2004); the two species cannot be distinguished on the basis of the size of the distal humerus. The two measurements taken on the distal humeri were the anterior breadth of the distal trochlea (DBt) and the minimum (antero-posterior) diameter of the (lateral-medial) center of the distal trochlea (MNd). Use of two measurements rather than one should make any effort to identify bilateral mates more accurate because specimens must be symmetrical in two dimensions rather than one.

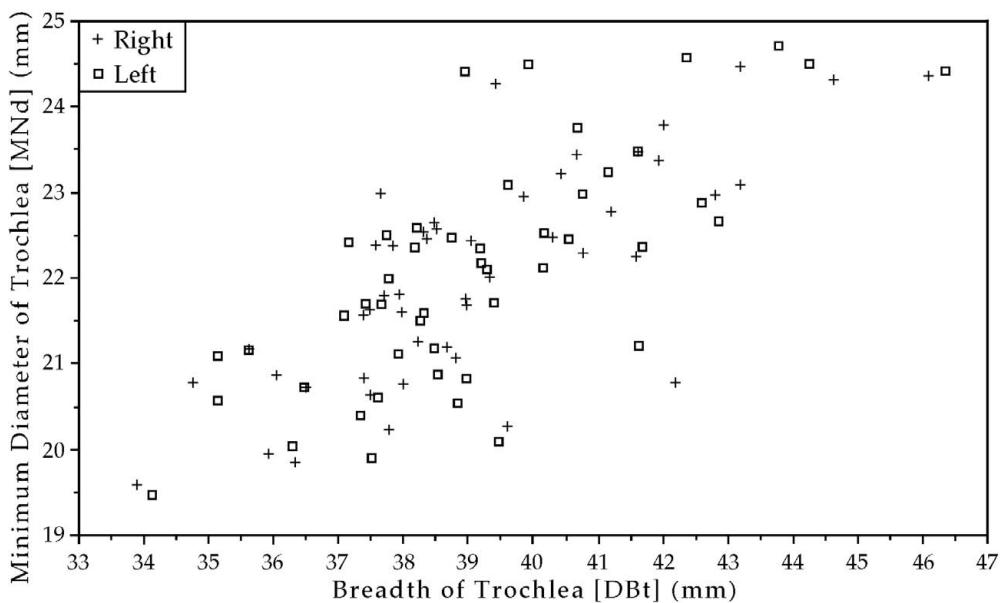
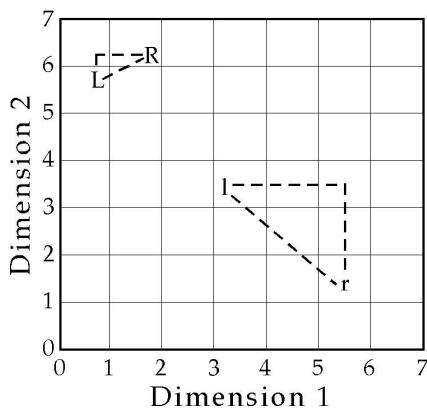
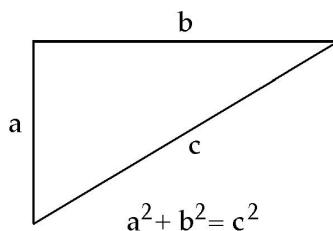


FIGURE 3.12. Latero-medial width of the distal condyle and minimum antero-posterior diameter of the middle groove of the distal condyle of forty-eight pairs of left and right distal humeri of *Odocoileus virginianus* and *Odocoileus hemionus*. Try to determine which left element is the bilateral match of which right element.

Conceive the difference between the left and right DBt as defining the length of one side of the right angle of a right triangle ( $= a$  in Figure 3.13), and the absolute difference between the left and right MNd as defining the length of the other side of the right angle of the right triangle ( $= b$  in Figure 3.13). Then, the Pythagorean theorem ( $a^2 + b^2 = c^2$ ) can be used to find how asymmetrical the distal humeri are in terms of the two measurements. If the left and right distal humeri were perfectly symmetrical, then the hypotenuse ( $c$ -value) of the right triangle would be zero because  $a = 0$  and  $b = 0$ . None of the distal humeri pairs include specimens that are perfectly symmetrical; the hypotenuse length or  $c$ -value of the triangle defined by the left and right specimens of each pair  $> 0.0$ . There is no statistically significant difference between the  $c$ -values displayed by pairs of distal humeri of the two deer species ( $O. virginianus$ ,  $0.753 \pm 0.81$ ;  $O. hemionus$ ,  $0.467 \pm 0.282$ ; Student's  $t = 1.771$ ;  $p > 0.08$ ), so the mean of all forty-seven, plus the hybrid (forty-eighth specimen), was determined; the average  $c$ -value =  $0.561 \pm 0.541$ . A tolerance level one might adopt is a  $c$ -value  $\leq 0.561$ ; any set of one left and one right distal humerus of deer the DBt and MNd measurements of which defined a  $c$ -value  $\leq 0.561$  would be identified as a bilateral pair.



**FIGURE 3.13.** A model of how two dimensions of a bone can be used to determine degree of (a)symmetry between bilaterally paired left and right elements. Upper, the Pythagorean theorem describing how the lengths of the three sides of a right triangle are related; lower, a model of how two measurements of a skeletal part considered together define a right triangle. The length of the hypotenuse ( $c$ -value in the text) is a measure of how (a)symmetrical two specimens are; specimens L and R are more symmetrical than specimens l and r.

Were you to try to identify bilateral pairs of one left distal humerus and one right distal humerus using the tolerance level just identified, you might get some of the pairs correct. But you likely could not match them all because of the chosen tolerance level (some pairs would have  $c$ -values  $> 0.561$ ), and of those that you did match, some would likely represent incorrectly identified pairs. With respect to the latter, consider Figure 3.12 again. Assuming we did not know which lefts went with which rights in this figure, use of the  $c$ -value of 0.561 to identify pairs in this data set is difficult if not impossible because of the large number of specimens. Earlier analysts seem to have recognized this problem, although their wording was inexplicit (e.g., Uerpmann 1973:311). Later commentators were more explicit. Nichol and Wild (1984:36–37) noted that “it is much harder to identify the extra unmatchable elements in a collection of bones from a larger number of animals of a species than it is in a smaller collection. For example, if there are 100 lefts and 100 rights of a bone, it will be rather difficult to produce an estimate of MNI much greater than 100, whether

100 or 200 animals are represented, while it may be very easy to see that a single left and a single right come from different individuals."

The blizzard of points in Figure 3.12 comprising both left and right specimens comprises the problem. Within the center of the distribution it is effectively impossible to determine which lefts go with which rights, and in a paleozoological sample approximating this size it would be impossible to identify true pairs and to find truly unpaired specimens. Only at the peripheries of this point scatter do we find individual left elements without multiple possible right mates and individual right elements without multiple possible left mates. Exacerbating the problem are the facts that populations of deer killed and deposited in many sites (= thanatocoenose) likely represent several dozen individuals, yet archaeologists generally collect only a sample of those remains – perhaps as little as 10 percent or even less – the Whitean MNI of the population (= thanatocoenose) might be 6–10. This is in part why the Lincoln–Petersen index seems so attractive – the manner in which it is calculated explicitly acknowledges that the materials at hand are a sample of a population (though which population in Figure 2.1 is unclear). But Figure 3.12 indicates that even were one to use the tolerance level or *c*-value of 0.561 for distal humeri based on the model in Figure 3.13, it is likely that false pairs would be identified. A more conservative *c*-value, such as, say, 0.25, would overlook many true pairs and still result in the identification of some false pairs.

Perhaps bilateral pairs could be identified using morphological criteria such as the size and rugosity of muscle scars and other traits, but deciding which left elements to compare with which right elements likely would begin with similarly sized specimens. Similarly, it might be possible to identify bilateral pairs in samples of fewer than a dozen lefts and fewer than a dozen rights, but if these specimens comprise but a sample of the population (and it is likely they will for one or both of two reasons – only part of the population was recovered because of sampling and recovery, or only part of the population was recovered because of differential preservation), then who is to say that the population would not approximate that in Figure 3.12 (see Lyman 2006a for a real example).

In light of this discussion of distal humeri and a similar analysis of deer astragali discussed elsewhere (Lyman 2006a), the assumption that analysts can accurately identify bilateral pairs required of quantitative units, such as the Lincoln–Petersen, index is unwarranted. And there is yet another seldom acknowledged difficulty with quantitative methods aimed at measuring or estimating taxonomic abundances that require information on the number of bilateral pairs in the assemblage.

The Lincoln–Petersen index is dependent on aggregation. Pairs would not be sought among a set of left skeletal elements from a stratum deposited about 3,000 years ago and a set of right elements recovered from a stratum deposited

about 2,000 years ago (unless perhaps one sought to measure postdepositional disturbance). This underscores in yet another way the potentially significant influence of aggregation on derived quantitative measures and estimates. Even if we have a single stratum, and we are comfortable with the procedure we use to identify bilateral pairs, have we gained any resolution with respect to taxonomic abundances? The data in Table 3.15 suggest it is unlikely that we have.

### CORRECTING FOR VARIOUS THINGS

When Shotwell (1955, 1958) discussed how he estimated taxonomic abundances, he noted that not only do individuals of different taxa have different numbers of skeletal elements per individual, individuals of different taxa also have different numbers of taxonomically identifiable skeletal elements. (For sake of simplicity, ignore the potentiality that skeletal elements might be broken and thus constitute specimens rather than skeletal elements.) Thus he advocated determination of the corrected number of skeletal elements per taxon, or “the number that would be expected if all species contributed the same number (the *standard number of elements*) of recognizable elements” (Shotwell 1955:331). To calculate the corrected number of skeletal elements (CNSE), he used the equation:

$$\text{CNSE} = [E(\text{SNE})]/\text{ENE},$$

where E is the number of elements identified in a collection, SNE is the standard number of elements per individual that the analyst can identify, and ENE is the estimated number of elements that the paleozoologist can identify in one complete skeleton of the taxon under consideration. SNE is, according to Shotwell (1955), an arbitrary number chosen so as to reduce the amount of corrective math.

It was argued in Chapter 2 that corrections for variation in the frequency of identifiable elements per taxon were unnecessary for several reasons. Nevertheless, such weighting of skeletal frequencies has been suggested by more than one paleozoologist (e.g., Gilinsky and Bennington 1994; Plug and Badenhorst 2006; Plug and Sampson 1996). Holtzman (1979), for example, suggests calculation of what he terms the “weighted abundance of elements” or WAE. This value is calculated as

$$\text{WAE} = \text{NE}_j/\text{NE}_{ij},$$

where  $\text{NE}_j$  is the frequency of skeletal elements of taxon  $j$  identified in an assemblage, and  $\text{NE}_{ij}$  is the frequency of skeletal elements of individual  $i$  of taxon  $j$  that can be identified. Because WAE uses the number of *skeletal elements* rather than the number of specimens, Holtzman (1979:80) argued that it “is less susceptible to biases arising

*Table 3.16. Abundances of beaver and deer remains at Cathlapotle, and WAE values and ratios of NISP and WAE values per taxon per assemblage*

	NISP	WAE	Ratio, beaver: deer
Beaver, precontact	111	1.247	NISP, precontact = 0.143
Beaver, postcontact	238	2.674	NISP, postcontact = 0.177
Deer, precontact	775	14.091	WAE, precontact = 0.088
Deer, postcontact	1,347	24.491	WAE, postcontact = 0.109

from variation in degree of fragmentation or number of [identifiable] elements per individual,” although he also acknowledged that it worked well only if differential preservation across taxa was not a problem. The methods used for determining the number of skeletal elements represented by a collection of anatomically incomplete specimens (an assemblage of broken bones) are considered in Chapter 4. It suffices here to note that Holtzman’s WAE can be described as the number of skeletal elements identified per taxon, adjusted to account for the number of elements in one complete skeleton that can be identified by the paleozoologist. An example will illustrate Shotwell’s and Holtzman’s concerns.

Although the frequencies of deer remains and of beaver remains from Cathlapotle are NISP (Table 1.3), let us treat them as if they are frequencies of anatomically complete skeletal elements. Frequencies of the two taxa are given in Table 3.16. If each tooth is considered as an independent skeletal element, the skull is considered one element, and vertebrae, carpal, and tarsals are ignored, each individual deer has the same number of skulls, mandibles, scapulae, humeri, radii, ulnae, innominate, femora, tibiae, and fibulae (distal end only in ungulates) as each individual beaver. But a beaver has two clavicles whereas a deer has none; a beaver has four incisors whereas a deer has eight; a beaver has sixteen metapodials whereas a deer has four; and a beaver has forty-eight phalanges (first, second, and third) whereas a deer has twenty-four. Thus a single deer can be conceived (for sake of discussion) to have fifty-five identifiable elements whereas an individual beaver has (for sake of discussion) eighty-nine identifiable elements.

If we follow Holtzman’s (1979) procedure, then (NISP) abundances of beaver and deer at Cathlapotle change to the WAE values indicated in Table 3.16. Notice that the relative abundances of beaver and deer do not change whether NISP or

WAE values are used. Deer always outnumber beaver. Notice as well that the WAE values provide a sort of fractional MNI in the sense that there are sufficient skeletal elements of beaver in the precontact assemblage to represent about one and a quarter individuals, and there are sufficient skeletal elements of beaver in the postcontact assemblage to represent a bit more than two and a half individuals. Calculation of Shotwell's CNSE or Holtzman's WAE does not gain us much accuracy in estimating taxonomic abundances.

Rogers (2000a) has been concerned with how the differential accumulation of skeletal parts in conjunction with the differential preservation of those parts might influence estimates of animal abundance (among other things). He notes that empirical work has focused on the two taphonomic processes of accumulation (what he terms deposition) and attrition, and that focus has resulted in major gains in knowledge. But he was concerned that statistical methods had not developed or improved at the same pace as empirical knowledge had improved (Rogers 2000b). Rogers therefore developed what he termed "analysis of bone counts by maximum likelihood" (ABCML). ABCML is a complex statistical algorithm into which one plugs various data such as skeletal part frequencies (see Chapter 6), bulk density per skeletal part (as a proxy for preservation potential), economic utility per part, and the like. Attrition of skeletal parts is assumed to be proportional to the bulk density of each part, and the probability of transport of a part is assumed to be proportional to the utility of a part. Rogers (2000a:122–123) explicitly acknowledged that ABCML was "incomplete" because it has empirical weaknesses. Weaknesses include the modeling of attrition and transport based on density and utility, respectively; actualistic (ethnoarchaeological) research indicates that both of these processes are influenced by more than just density and utility. Application of the ABCML protocol to a zooarchaeological collection (Rogers and Broughton 2001) makes these points explicitly clear for two reasons. First, the assumptions that are required are numerous, and second, the qualitative results are characterized as "more trustworthy" and more likely to be correct than are the quantitative results (Rogers and Broughton 2001:763, 772).

It is likely that ABCML has not often been used for two reasons. First, many paleozoologists lack the statistical sophistication necessary to comprehend what ABCML is and how it works (Rogers 2000b). Paleozoologists must learn more about statistical methods, and they must overcome about seven decades of disciplinary historical inertia that has focused on deterministic questions rather than probabilistic ones (Lyman 1994c). Second, the method requires one to assume much regarding the taphonomic history of the collection. Has the collection been subjected to differential transport intrataxonomically or intertaxonomically or both? Has it been subjected to differential attrition intrataxonomically or intertaxonomically or both? A colleague

suggested that other methods analytically overlook these possibilities when estimating taxonomic abundances, but ABCML deals with them. It also provides confidence intervals, allowing one to assess how tight (or loose) an estimate is.

Rogers and Broughton (2001) advocate the use of ABCML because of perceived flaws in more simplistic analyses. In particular, they correctly note that even non-parametric measures of association such as Spearman's rho (between, say, NISP and MNI) assume that NISP counts represent independent tallies. We know NISP counts do not (or are highly unlikely to) represent tallies of independent specimens (Chapter 2). Correlation coefficients also assume that different taxa have equal numbers of identifiable skeletal elements (ignoring fragmentation), and we know that they do not. Finally, Rogers and Broughton (2001) correctly argue that different values of a correlation coefficient cannot be related theoretically to the intensity of a taphonomic process. Correlation coefficients are, however, used by paleozoologists to gain insight to possible associations; none builds an argument on just a correlation coefficient and instead consults other pertinent data to help understand why a correlation exists (or doesn't exist). Few paleozoologists infer the intensity or degree of a taphonomic process simply on the basis of the magnitude of a correlation coefficient; rather, a coefficient is usually interpreted in nominal-scale terms. Two variables are correlated, or they are not. What the presence (or absence) of a correlation means is a taphonomic issue more than a statistical one.

ABCML presents a detailed view of the sorts of variables that one must consider in any analysis of transport and attrition and how those processes are likely to influence taxonomic abundances. In this I (in Lyman 2004c) agree completely with Rogers's (2000a:123) observation that detailed consideration of the empirical requirements of ABCML will "help identify the [taphonomic, biological, archaeological, etc.] parameters that should be estimated and reported." Time will tell if the method gains popularity among paleozoologists.

## SIZE

In a recent discussion, Orchard (2005) argued that the relationship of bone size to body size can be used to assist with the estimation of taxonomic abundances. The method is easy to grasp conceptually. Each skeletal element has a particular statistical relationship to body size that can be established with museum specimens. That relationship can be described by a regression equation. Once such equations are established for multiple skeletal elements, different skeletal elements in a prehistoric collection may be used to estimate the body sizes represented. Let's say that the most

common skeletal element in a prehistoric collection is right astragali, and the sizes of those specimens suggest that there are five individuals (= MNI) of varied sizes. If distal right tibiae suggest that there are only four individuals (= MNI) but one of those tibiae indicates an individual body size larger than any of the individual body sizes indicated by right astragali, then a sixth individual is added to the tally of individuals.

There is a practical problem; “the time and effort involved in gathering comparative data and generating regression formulae, as well as the difficulty in obtaining adequate comparative samples, can be prohibitive” (Orchard 2005:357). Generating regression formulae is relatively easy. It is the process of acquiring the requisite data that is difficult. Consider, for example, Emerson’s (1978) data for white-tailed deer summarized in Table 3.12. He worked the several weeks of an annual hunting season. I spent 2 weeks visiting eight collections of deer skeletons in various museums and comparative zooarchaeological collections to generate the data presented in Figure 3.12. Those collections are housed in widely separated localities (Wyoming, Montana, Washington, British Columbia). There is a more fundamental problem with Orchard’s suggested procedure, however, and it can be illustrated with some of the data collected when the Figure 3.12 data were collected.

Consider Emerson’s (1978) data summarized in Table 3.12. As noted earlier, those data provide the equation  $Y = -104.96 + 4.11 X$ , where  $Y$  is the body weight or individual biomass in kilograms, and  $X$  is the maximum lateral length of the astragalus in millimeters. Applying that equation to the seventeen left and seventeen right astragali constituting bilaterally paired elements of white-tailed deer that I measured produces the results summarized in Table 3.17. Variation between the individual body size estimated by the length of the left astragali and the body size estimated based on the length of its bilaterally paired right astragali mate ranges from 0.25 kilograms to 2.3 kilograms with an average of 0.94 kilograms. Similar analysis of forty-three pairs of astragali from mule deer indicates difference in body size estimates provided by left and right elements averages 1.01 kilograms and ranges from 0.08 to 4.11 kilograms. The problem that presents itself is precisely the one illustrated in Figure 3.13 but the variables are different. In the former case the problem concerned the degree of symmetry of distal left and right humeri in terms of size; now it is symmetry in estimates of body weight derived from the size of astragali. Recall that only those specimens that provide asymmetrical results (do not have bilateral mates in the collection) will also add to the tally of individuals represented by an assemblage of skeletal elements. What tolerance level should be chosen and why? How symmetrical should the two estimates of body size be in order to conclude the size of the same individual has been estimated twice? Orchard (2005) provides no guidance, and he is

Table 3.17. Estimates of individual body size (biomass) of seventeen white-tailed deer based on the maximum length of right and left astragali. Estimation equation is  $Y = -104.96 + 4.11X$ , where  $Y$  is the body weight or individual biomass in kilograms, and  $X$  is the maximum lateral length of the astragalus in millimeters (after Emerson 1978)

Pair	Length of right	Right weight	Length of left	Left weight	Difference
1	38.78	54.426	38.36	52.700	1.726
2	40.38	61.002	40.08	59.769	1.233
3	43.10	72.181	42.86	71.195	0.986
4	39.90	59.029	40.00	59.440	0.411
5	36.84	46.452	36.74	46.041	0.411
6	43.74	74.811	43.32	73.085	1.726
7	38.94	55.083	38.88	54.837	0.246
8	42.76	70.784	42.44	69.468	1.316
9	38.94	55.083	38.38	52.782	2.301
10	43.74	74.811	43.46	73.661	1.150
11	42.86	71.195	42.94	71.523	0.328
12	41.72	66.509	41.36	65.030	1.479
13	41.80	66.838	41.92	67.331	0.493
14	37.44	48.918	37.54	49.329	0.411
15	39.70	58.207	39.82	58.700	0.493
16	40.88	63.057	40.72	62.399	0.658
17	40.52	61.577	40.38	61.002	0.575

wise to not do so (Lyman 2006a). All of the problems that attend identifying bilateral pairs also plague Orchard's (2005) method.

## DISCUSSION

None of the quantitative units and methods occasionally used to estimate or measure taxonomic abundances reviewed in this chapter have been widely adopted or seen more than sporadic use. When Grayson (1984) wrote his synopsis of quantitative zooarchaeology, he focused on NISP and MNI because they were at the time the most widely used units. His discussions of other methods such as the Lincoln–Petersen index were terse. I have attempted here to empirically support, or refute claims regarding these other methods. Thus, we find that biomass and meat weight

estimates compound many weaknesses of MNI because of requisite assumptions regarding average live weights and edible tissue amounts. Ubiquity as a measure of the “importance” of a taxon is strongly influenced by sample size measured as NISP; it provides information on taxonomic abundance that is virtually identical to information provided by NISP. Ubiquity might measure some as yet unknown (taphonomic?) variable if two taxa with statistically indistinguishable sample sizes have different ubiquities, but what that variable might be is unclear.

Numerous quantitative methods have been proposed as improvements to MNI. Virtually all involve investing what Theodore White thought would be a great deal of time determining which left elements pair up bilaterally with which right elements. The validity of estimates of taxonomic abundance provided by those methods rests on the validity of pair identification. The pair identification procedure rests on the notion of bilateral symmetry but no organism is perfectly bilaterally symmetrical, so one must decide how symmetrical is symmetrical enough. Even if that decision can be made, in large samples (of more than, say, two dozen specimens) potential bilateral mates for any specimen are multiple. This highlights the fact that false pairs are likely to be identified even in small samples of lefts and rights.

Other methods discussed in this chapter concern efforts to correct for intertaxonomic variation in (i) number of identifiable elements per individual, (ii) transport or accumulation, and (iii) fragmentation. These variables can be analytically accounted for in NISP (see Chapter 2). Where are we left, then, with respect to measuring taxonomic abundances? As alluded to in the preceding chapter, I agree with Grayson (1979, 1984) and the numerous paleobiologists who use NISP to measure taxonomic abundances. It is cumulative or simply additive, meaning it is primary data or an observed measure; and it serves as the basis for many derived measures (and hence is correlated with many of them). I use it virtually exclusively in subsequent chapters when issues of taxonomic abundance are under study. When I do not, I explain why; generally I use quantitative units other than NISP when the target variable is not taxonomic abundances.

# Sampling, Recovery, and Sample Size

It is commonsensical to note that what is recovered – the *amount* recovered of each kind, and the number of kinds – will influence quantitative analyses (Cannon 1999). As we have seen in earlier chapters, sample size influences many measures and estimates of taxonomic abundances. The size of a sample of faunal remains measured as the number of specimens recovered is in turn influenced by the sampling design chosen (how much is excavated) and the recovery techniques (passing sediment through fine- or coarse-mesh sieves) used to implement that sampling design. This chapter focuses on how one generates a collection of faunal remains (sampling and recovery), properties of the resultant sample, and ways to examine the influences of sample size on selected target variables.

Paleozoologists have long worried about how methods of recovery might produce collections that are not representative of a target variable (e.g., Hibbard 1949; Kuehne 1971; McKenna 1962; Payne 1972; Thomas 1969). Exacerbating this worry is the fact that paleozoologists collect *samples* of faunal remains from geological contexts (Krumbein 1965; Ward 1984). This is true on at least two levels. First, paleozoologists never (or at least very seldom) collect all of the faunal remains from a deposit, paleontological location, or archaeological site. Second, the target variable usually resides in an entity other than the “identified assemblage” (Figure 2.1). If the target variable is the taphocoenose (either that which is preserved or that which was deposited), the thanatocoenose, or the biocoenose, the paleozoologist is dealing with a sample regardless of whether or not the complete deposit has been excavated.

For more than 50 years, paleozoologists have suggested that probabilistic sampling methods will produce “representative samples” (e.g., Gamble 1978; Krumbein 1965; Voorhies 1970). These methods concern techniques to choose portions of the geological record to examine for faunal remains. There are many excellent discussions

of probabilistic sampling (e.g., Orton 2000), so methods of probabilistic sampling are not discussed here. Instead the focus is on a couple of related sampling issues. Choosing deposits to inspect for faunal remains is but one part of the *sampling problem*. Another part concerns how faunal remains are retrieved or collected from deposits chosen for inspection. If more sediment samples are chosen for inspection, or more remains are recovered from the chosen sediments because of how remains are retrieved, then the resultant sample is different (larger) than it might otherwise have been.

In this chapter, two issues are of concern. One is collection or recovery technique. Are faunal remains picked by hand from exposed sediments; are sediments passed through hardware cloth (screens or sieves) and faunal remains that do not pass through the cloth collected; are faunal remains collected from bulk samples, from flotation samples, or by some other means? Because the recovery methods used influence what is collected, efforts have been made to correct for these influences, and some of these correction procedures are reviewed here. Another issue discussed in this chapter is *sample size* measured in either or both of two ways – as amount of sediment examined (either area or volume) and as amount of faunal material studied. Both measures of sample size often correlate with the number of taxa recovered, the relative abundances of those taxa, and the like. To make valid interpretations of quantitative faunal data, we must understand both the nature of a sample and how sampling techniques may influence the faunal variables that we hope to measure. All of the variables discussed here – amount of sediment inspected, screen mesh size, NISP – are particular manifestations of *sampling effort*. Greater sampling effort (however measured) will produce larger samples, but how sampling effort influences other characteristics of the sample is not always recognized.

One distinction that must be made at the start concerns the difference between *discovery sampling* and *statistical precision sampling* (Nance 1983). Discovery sampling concerns sampling designs built to discover new phenomena and the sampling effort required to find various categories of phenomena. The more rare a kind of thing is in a sampling universe, the more sampling effort required to find an instance of that kind. Statistical precision sampling generates a sample that provides an accurate estimate of a target variable. Whereas discovery sampling focuses on finding examples of rarely occurring kinds of phenomena, statistical precision sampling seeks to estimate properties of commonly occurring categories of phenomena, whether abundances of individual instances within each category, average size of members of each category, or any of a plethora of other variables that might be measured. The distinction of discovery sampling and statistical precision sampling will be important in this and subsequent chapters.

## SAMPLING TO REDUNDANCY

Many of the quantitative variables that we seek to measure with paleozoological collections are often strongly influenced by sampling effort (how ever such effort is measured), itself a quantitative variable. In this chapter analytical techniques that have been suggested for controlling those influences when comparing samples of markedly different sizes are outlined. These techniques are based on the assumption that no other deposits will be inspected for faunal remains, and thus that no new faunal remains will be added to the samples at hand. Another method that assumes new specimens are forthcoming until a sample that is representative of the target variable(s) is in hand is also described. This latter method can be implemented with either or both of two distinct measures of sample size.

Paleozoologists typically collect a sample of faunal remains from a population of remains. (For sake of discussion, the identity of the target variable – taphocoenose, thanatocoenose, biocoenose – will be ignored.) The population may comprise all the faunal remains encased within a stratum, or those within several strata thought for non-faunal reasons to represent the same zoological property of interest and thus for analytical purposes to be instances of the same population. Because paleozoologists sample the depositional (geological) record for faunal remains, they generally collect multiple samples. One sample may be collected today and another tomorrow; one sample may be collected from a particular geographic and geological location and another from a different location. In many cases, individual samples are collected over multiple time periods, whether those periods are consecutive months or consecutive annual field seasons. Because consecutively gathered samples from a deposit (from what is thought to represent the same population) are cumulative, a basic method of empirically assessing sample adequacy suggests itself.

Assume that a target variable has been specified by the research problem one is attempting to solve. Let us say that the target variable requires measurement of the number of mammalian taxa in a collection, generally known as taxonomic richness. The acronym NTAXA for “number of taxa” will be used here. (NTAXA is used by ecologists and archaeologists [e.g., Broughton and Grayson 1993] to measure niche breadth [among other things].) How do we know when we have collected enough faunal remains to have a sample that provides a relatively accurate estimate of NTAXA? Archaeologist Robert Leonard (1987:499) suggested that one could sample “to redundancy” and that the way to know when additional samples were redundant with previous samples was simple; “plot the information gained against the number of samples taken and determine if the curve is becoming asymptotic. It may then be reasonable to assume that the sample is sufficiently representative with regard to that

Table 4.1. *Volume excavated and NISP of mammals per annual field season at the Meier site*

Year	Volume (m <sup>3</sup> )	NISP	NTAXA	Deer NISP	Deer NISP/m <sup>3</sup>
1973	11.0	519	16	276	25.1
1987	40.7	1,359	21	778	19.1
1988	31.2	1,232	22	756	24.2
1989	46.3	970	19	562	12.1
1990	29.2	956	18	570	19.5
1991	37.7	1,385	20	838	22.2

particular information.” In paleozoology, sample size can be measured one of two ways – either as amount excavated or as NISP.

#### *Excavation Amount*

Paleontologists have examined the influence of sample size measured as amount of sediment examined on NTAXA (e.g., Raup 1972). Efforts to correct for such continue to this day (e.g., Crampton et al. 2003). Wolff (1975) used an empirical means to determine if sufficient sediment had been examined to argue that his samples were representative of NTAXA. He compiled data on cumulative NTAXA across increasing amounts of sediment from which faunal remains had been extracted. When his cumulative NTAXA curve leveled off across several additional units of sediment volume, Wolff (1975) argued that his total sample was representative of that target variable. With the cumulative NTAXA plotted on the vertical or y-axis of a bivariate plot, and new unit volumes of sediment added along the horizontal or x-axis of the plot, Wolff showed that taxa were initially added quickly by new samples, but as the number of samples increased the rate of addition of new taxa slowed until it leveled off across multiple new samples. The latter was taken by Wolff to mean that he had sampled sufficiently to have a representative sample; Leonard (1987) would say that Wolff had sampled to redundancy because faunal remains from additional unit volumes of sediment failed to produce any new taxa.

The protocol is easy to illustrate with the data in Tables 4.1 and 4.2 for the Meier site. If we use those data to construct a cumulative NTAXA curve based on volume excavated, we obtain the result in Figure 4.1. As the volume excavated from one year to the next increased, NTAXA initially increased, but then it leveled off and no

Table 4.2. Annual NISP samples of mammalian genera at the Meier site

Taxon	1973	1987	1988	1989	1990	1991	Total
<i>Scapanus</i>	4	4	3	4	1	2	18
<i>Sylvilagus</i>	2	3	1	1	10	1	18
<i>Aplodontia</i>	2	1		1		3	7
<i>Tamias</i>				1			1
<i>Tamiasciurus</i>			2				2
<i>Thomomys</i>		2		1	5	1	9
<i>Castor</i>	13	100	65	52	41	71	342
<i>Peromyscus</i>		4	12	12	4	3	35
<i>Rattus</i>			1				1
<i>Neotoma</i>			1				1
<i>Microtus</i>		15	25	34	15	11	100
<i>Ondatra</i>	37	97	55	59	74	52	374
<i>Erethizon</i>				1			1
<i>Canis</i>	2	25	13	16	11	25	111
<i>Vulpes</i>	3	1	1				5
<i>Ursus</i>	20	16	20	7	13	26	102
<i>Procyon</i>	15	79	51	35	43	64	287
<i>Martes</i>	1	6	1		1	11	20
<i>Mustela</i>	4	35	17	19	38	21	134
<i>Mephitis</i>		1	1			2	4
<i>Lutra</i>	6	12	6	2	11	14	51
<i>Puma</i>		4	1		3	1	9
<i>Lynx</i>	9	5	4	1	4	8	31
<i>Phoca</i>	3	6	5	10	6	13	43
<i>Cervus</i>	103	165	191	152	106	218	935
<i>Odocoileus</i>	276	788	756	562	570	838	3,780
Annual $\sum$ NISP	519	1,359	1,232	970	956	1,385	6,421
Annual NTAXA	16	21	22	19	18	20	26
Cumulative $\sum$ NISP	519	1,878	3,110	4,080	5,036	6,421	—
Cumulative $\sum$ NTAXA	16	21	24	26	26	26	—

new taxa were added after the first 129.2 m<sup>3</sup> of sediment had been inspected. With respect to cumulative volume of sediment excavated, the Meier site sample contains representatives of at least the most common taxa (see the following section); very rarely represented taxa may not be present in the collection, but the sampling to redundancy procedure suggests that we have a statistically precise representation of the common taxa.

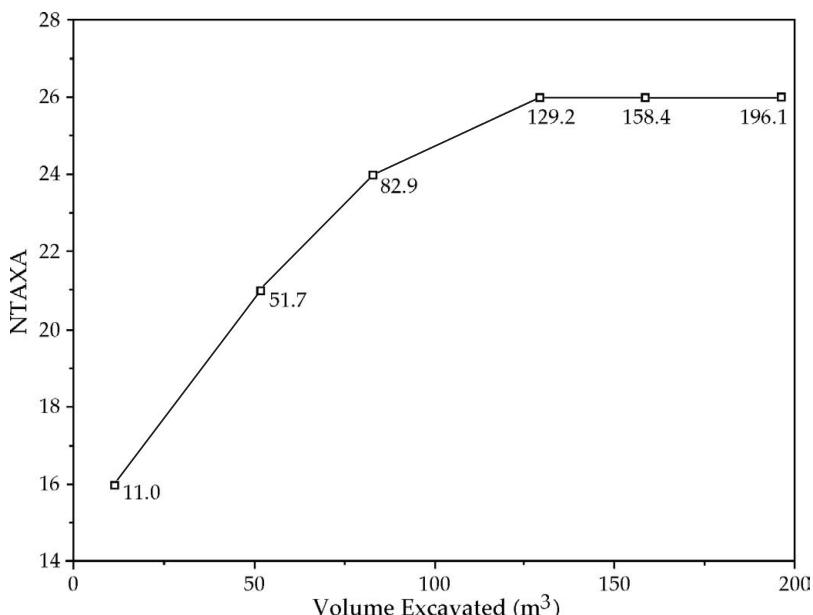


FIGURE 4.1. Cumulative richness of mammalian genera across cumulative volume (m<sup>3</sup>) of sediment excavated annually at the Meier site. Numbers adjacent to plotted points are cumulative m<sup>3</sup>. Data from Table 4.1.

#### *NISP as a Measure of Sample Redundancy*

Retaining taxonomic richness or NTAXA as our target variable for illustrative purposes, consider the Meier and the Cathlapotle collections. Meier was sampled over a period of six annual field seasons (the last five were consecutive) by two archaeologists. Cathlapotle was sampled over a period of four annual field seasons by one archaeologist, but early in the first annual field season recovery techniques varied considerably from those used later that year, so the first year is split into two chronologically consecutive samples for illustrative purposes. At both sites, each annual field season spanned a period of 8 weeks. Annual NISP samples from Meier are described in Table 4.2 and those for Cathlapotle are described in Table 4.3. Summed values for Meier in Table 4.2 are larger than those given in Table 1.3 because an additional sample analyzed in 1973 is included in the former table. Values for Cathlapotle in Table 4.3 are greater than those given Table 1.3 because included in the former table are specimens that could not be assigned to a temporal component and thus could not be included in Table 1.3.

It has long been recognized that the order in which samples are added to cumulative frequency curves can influence the result (e.g., Kerrich and Clarke 1967). The total

Table 4.3. Annual NISP samples of mammalian genera at Cathlapotle. The two 1993 samples represent different recovery techniques

Taxon	1993a	1993b	1994	1995	1996	Total
<i>Didelphis</i>					10	10
<i>Scapanus</i>					3	3
<i>Sorex</i>			4			4
<i>Lepus</i>			14	20	18	52
<i>Aplodontia</i>	2	18	41	42	33	136
<i>Castor</i>	1	32	123	185	51	392
<i>Peromyscus</i>			4	1		5
<i>Microtus</i>	1		12	16	39	68
<i>Ondatra</i>		19	34	32	21	106
<i>Canis</i>		4	27	5	3	39
<i>Vulpes</i>		1	3	1		5
<i>Ursus</i>	1	23	29	31	18	102
<i>Procyon</i>	1	57	59	70	20	207
<i>Martes</i>				2		2
<i>Mustela</i>		3	14	7	5	29
<i>Mephitis</i>					3	3
<i>Lutra</i>		14	19	13	19	65
<i>Puma</i>		5	3	3	1	12
<i>Lynx</i>		2	6	12	6	26
<i>Phoca</i>		1	19	41	4	65
<i>Ovis</i>			2			2
<i>Cervus</i>	16	462	879	1,184	683	3,224
<i>Odocoileus</i>	18	332	797	821	408	2,376
<i>Equus</i>			2	2		4
Annual $\sum$ NISP	40	973	2,091	2,488	1,345	6,937
Annual $\sum$ NTAXA	7	14	20	19	18	24
Cumulative $\sum$ NISP	40	1,013	3,104	5,592	6,937	—
Cumulative $\sum$ NTAXA	7	15	20	21	24	—

cumulative sample size at which the curve levels off and thus suggests that new samples are providing no new information but instead only redundant information can vary considerably depending on the order of sample addition. Thus, choice of the order in which samples are added must be explicit and logical. Given that there is an inherent (chronological) order to the annual samples from Meier and also to those from Cathlapotle, it is logical to treat the samples as cumulative in the temporal order in which they were collected. Doing so for the Meier annual samples produces the

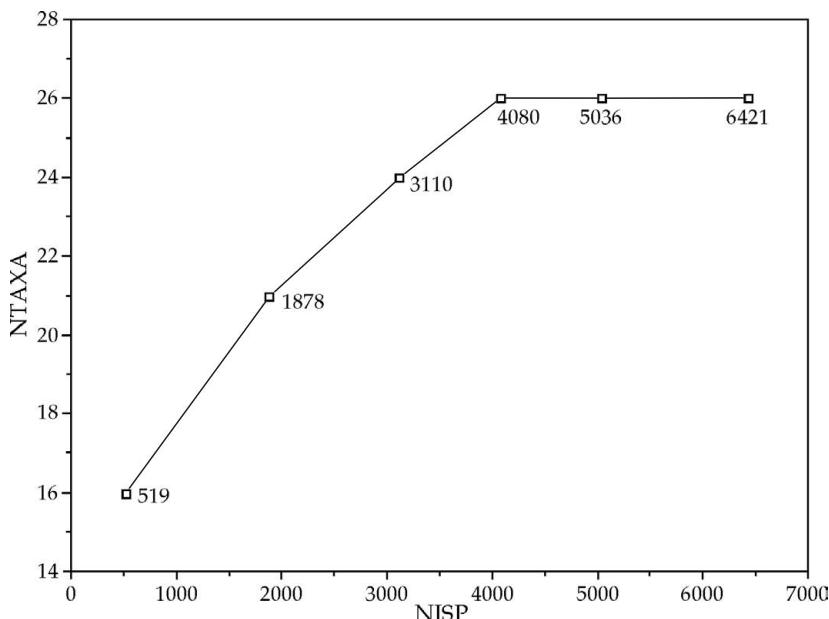
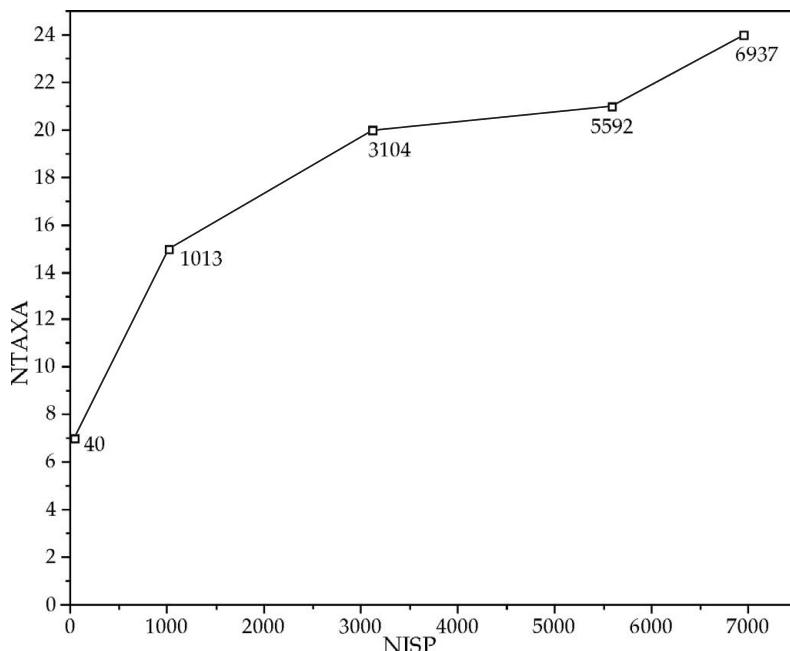


FIGURE 4.2. Cumulative richness of mammalian genera across cumulative annual samples (NISP) from the Meier site. Numbers adjacent to plotted points are cumulative NISP. Data from Table 4.2.

cumulative NTAXA curve shown in Figure 4.2; doing so for the Cathlapotle annual samples produces the cumulative NTAXA curve shown in Figure 4.3 (both curves are slightly different than those described in Lyman and Ames [2004] because all taxa are included here; Lyman and Ames [2004] excluded historically introduced taxa). What do those curves suggest?

On the one hand, the cumulative NTAXA curve for Meier levels off after the addition of the fourth, or 1989, sample (Figure 4.2). Despite an addition of more than 2000 NISP, no new taxa are added with the 1990 and 1991 samples. These last two, most recent samples are redundant with earlier samples in terms of their influence on the target variable of NTAXA. This suggests that the total Meier collection can be treated as representative of the mammalian genera deposited at the site. The cumulative NTAXA curve for the Cathlapotle sample, on the other hand, does not level off but rises with the addition of each new sample (Figure 4.3). An argument cannot be made that additional collection of faunal remains from this site will fail to produce evidence of additional mammalian genera. The cumulative NTAXA curve for Cathlapotle also suggests that the total sample, though it consists of nearly 7,000 NISP, does not represent all mammalian genera in the site deposits.

In the ecological literature, curves such as those illustrated in Figures 4.1, 4.2 and 4.3 are sometimes referred to as “accumulation curves” (Gotelli and Colwell 2001)



**FIGURE 4.3.** Cumulative richness of mammalian genera across cumulative annual samples (NISP) from Cathlapotle. Numbers adjacent to plotted points are cumulative NISP. Data from Table 4.3.

for an obvious reason. Sampling to redundancy has not been mentioned very often in paleozoological research (e.g., Lyman 1995a; Monks 2000; Reitz and Wing 1999:107), and used even less often (e.g., Butler 1990; Lyman and Ames 2004; Wolff 1975). Many paleoethnobotanical examples are cases in which sampling effort is plotted against richness, and the influences of sample size differences are noted (Lepofsky and Lertzman 2005). This underscores the ease with which a quantitative tool can be misrepresented as doing one thing when in fact it is doing something else. We return to this general kind of curve later in this chapter. Here it suffices to note that the curves in Figures 4.1, 4.2 and 4.3 are but one kind of a more general kind of curve that is used to examine the relationship between sample size and ecological variables such as NTAXA.

#### *Volume Excavated or NISP*

The amount of sediment examined for faunal remains is one measure of sample size, but the NISP per unit volume of sediment can vary considerably. This means that correlations between sediment volume and, say, NTAXA, are likely to be less strong

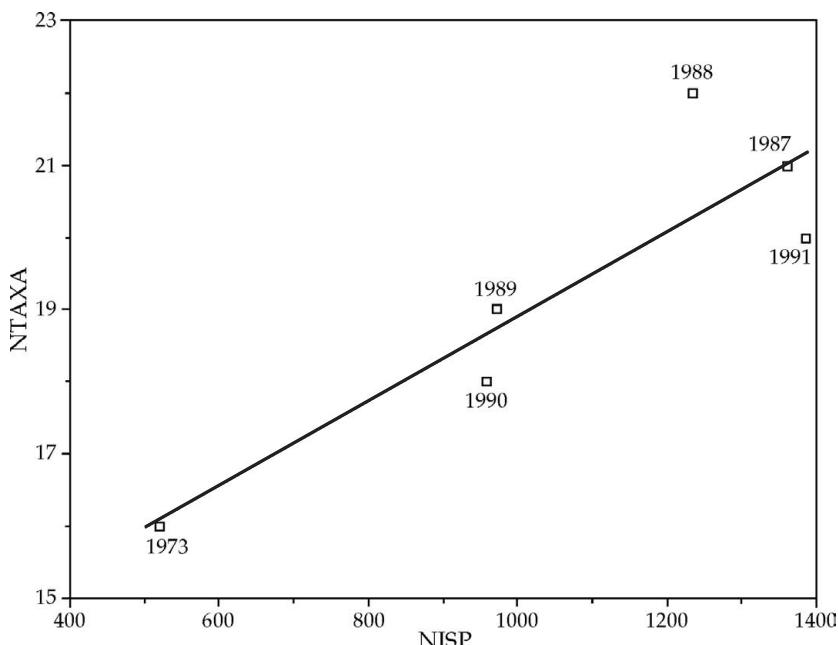


FIGURE 4.4. Relationship of mammalian genera richness (NTAXA) and sample size (NISP) per annual sample at the Meier site. The relationship is described by the simple best-fit regression line ( $Y = 13.048 + 0.0059X$ ) and is significant ( $r = 0.89, p < 0.01$ ). The year the sample was collected is indicated. Data from Table 4.2.

that those between NISP and NTAXA. The NISP of deer varies by as much as thirteen NISP per cubic meter across the six samples from the Meier site (Table 4.1). There is no statistically significant relationship between the volume excavated per year and the NISP per annual sample for the 1987–1991 samples ( $r = 0.05, p > 0.2$ ); the smallest (1973) sample is deleted because it influences the result considerably ( $r = 0.73$  if that sample is included). NISP per annual sample and richness per annual sample at the Meier site, on the other hand, are strongly correlated (Figure 4.4). This suggests that the better variable with which to monitor the influence of sample size on variables such as NTAXA is NISP, though this may vary.

Were I to examine the adequacy of the sample from Cathlapotle in a real analysis rather than simply illustrating an analytical technique, I would apply the sampling to redundancy protocol to the precontact assemblage and also to the postcontact assemblage rather than to the site collection as a whole. This protocol demands that the target variable of interest be explicitly defined and the boundaries of the appropriate sample be unambiguous. Nevertheless, several lessons can be taken from the preceding. First, the absolute size of a sample is not necessarily a good measure of that sample's representativeness of a particular target variable. The total mammalian

genera NISP from Cathlapotle (= 6,937) is larger than the total mammalian genera NISP from Meier (= 6,421), yet the latter seems to be representative of NTAXA whereas the former does not seem to be representative of NTAXA. Second, cumulative chronological samples, whether by week, month, or year, provide logical units with an inherent cumulative order that may provide an indication of when enough material has been collected. Such an argument presumes that identification of the recovered faunal remains proceeds apace with recovery, or that the time lag between the two is insignificant. If identification can keep pace with recovery, then paleobiological resources can be saved in situ rather than disturbed (some would say “destroyed”) by recovery because it will be clear when a sample sufficiently large to provide an accurate answer (one not influenced by inadequate sample size) has been collected.

The final lesson is that, presuming one knows the identity of the variables plotted on both axes (and there is no reason the analyst should not), the meaning of the cumulative curve is commonsensical. When the curve is steep, much new information is being added with each new sample; when it is horizontal, new samples are adding no new information about the target variable. This makes such a curve useful as a simple (and readily visible) way to monitor what is being learned as sample size increases, and to determine whether additional samples are necessary or not. Samples can comprise the material collected during one temporal period, or they can be structured some other way, such as choosing (using probability sampling) a sample of 10 percent of all units excavated or exposures inspected (or collected), then another 10 percent, then another, and so on. Similarly, any target variable that consists of a single value can be plotted on the y-axis, just as any kind of sample can be plotted on the x-axis. Such target variables might involve the average or mean size of individuals of a taxon, or the frequencies of skeletal parts of a taxon, or virtually any variable.

There is more to say about the type of graph shown in Figures 4.1–4.4. This graph type is one in which a measure of sample size is plotted against a measure of a biological property, and more is said about such graphs and different versions of them later. The sampling to redundancy type of graph is introduced here to illustrate its value for evaluating sample adequacy as sample size is being actively increased. Once the fieldwork is completed, little else can be done to increase the size of a collection. Knowing whether more specimens are needed as collection is taking place would be valuable knowledge. Knowing whether one is losing material rather than collecting it as sediment is inspected (e.g., screened) would be equally valuable knowledge. Zooarchaeologists in particular have devoted a great deal of energy to figuring out how to generate this latter sort of knowledge, and it is to the results of those energy expenditures that we turn to next.

## THE INFLUENCES OF RECOVERY TECHNIQUES

Once a geographic and geological context in which to look for faunal remains has been chosen, the next step is to choose how those remains will be searched for and retrieved from sediments. Faunal remains can be *hand picked* from sediments as those sediments are excavated. Bones and teeth can be collected from screens or sieves the function of which is to allow sediment to pass through whereas faunal remains are caught in the mesh where they are more visible than when in the sediment in the excavation. Screens were not always used by zooarchaeologists who gathered, by hand, those bones and teeth they saw in the sediment as it was excavated. Watson (1972) showed that many bone fragments  $\leq 3$  cm maximum dimension tended to be overlooked when hand picking alone was used. Passing sediments through screens increased the return of small fragments an order of magnitude. An earlier study showed exactly the same thing using remains of mollusks.

### *Hand Picking Specimens by Eye*

Sparks (1961) demonstrated that the percentage of recovered remains of terrestrial mollusks differed markedly by size class. The sample collected by eye, unaided by screens, tended to have more specimens representing large size classes (>50 percent of all specimens recovered) whereas the sample collected from a screen was dominated by specimens representing small size classes (>70 percent of all specimens recovered). His data are graphed in a way different than Sparks did in Figure 4.5 to allow comparison of the taxonomic abundances in the two samples. The identity of the taxa themselves is unimportant to this exercise, so categories of specimens are distinguished on the basis of ordinal scale average size. Thus, whereas Sparks (1961) distinguished eighteen taxa, there are only fifteen size classes here. Figure 4.5 shows that the taxa with the largest shells were those that were most often collected by hand, and those taxa with the smallest shells were seldom collected by hand. What seems to have been a 2 mm mesh sieve produced many more individuals of small size, and Sparks (1961:72) concluded in an understated way that “Any attempt to pick out shells by eye from a deposit is bound to lead to distortion in the percentage frequencies of species.” Study by invertebrate paleobiologists of what is now referred to as *size bias* continues (e.g., Cooper et al. 2006; Kowalewski and Hoffmeister 2003).

Results like those Sparks (1961) derived for mollusks were found by Payne (1972, 1975) for mammal remains. Although he did not explicitly list the body size or average live weight of an adult animal, Payne (1975) found that more of the larger remains

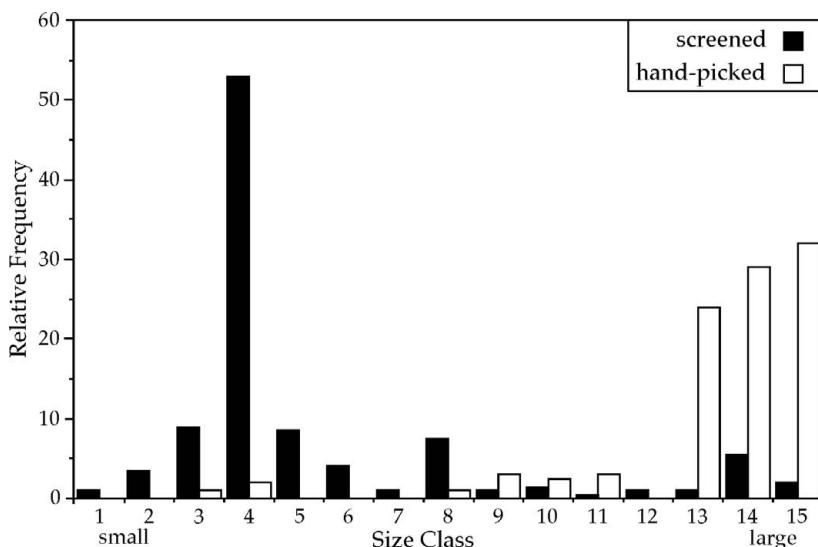


FIGURE 4.5. Relative abundances of fifteen size classes of mollusk shells recovered during hand picking from the excavation, and recovered from fine-mesh sieves. Original data from Sparks (1961).

of large-bodied taxa were found by hand while excavating whereas more of the small remains of small-bodied taxa were found in sieves or screens. Taxa were rank ordered in five size classes, from largest to smallest: cattle (*Bos* sp.), pig (*Sus* sp.), sheep and goat (*Ovis* sp., *Capra* sp.), canid (*Canis* sp., *Vulpes* sp.), and hares (*Lepus* sp.). Assuming that the remains that were recovered by hand picking from the excavation would also be recovered from the screen. Figure 4.6 shows two things about Payne's (1975)

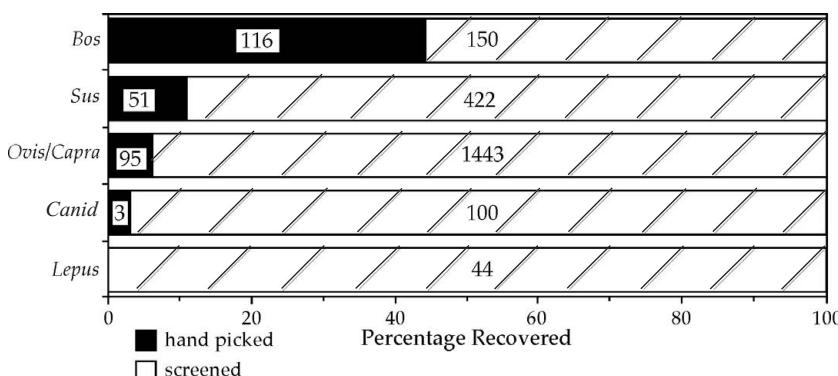


FIGURE 4.6. The effect of passing sediment through screens or sieves on recovery of mammal remains relative to hand picking specimens from an excavation unit. Numbers within bars are NISP. Data from Payne (1975).

data. First, more remains in general are collected from screens than by hand from an excavation, something not so obvious given how Sparks (1961) presented his data. Second, the smaller the body size of a taxon, the more of its remains will be found in the screen than in the excavation; this echoed Sparks's original observation on mollusk remains but expanded it to include remains of mammals.

### *Screen Mesh Size*

It is commonsensical to believe that small bones and small fragments thereof will fall through coarse-mesh hardware cloth (that with large holes) whereas many will be caught by and thus be recovered if fine-mesh hardware cloth is used. Thomas (1969) and Payne (1972, 1975) demonstrated this empirically (see also Casteel 1972; Clason and Prummel 1977), and showed that the magnitude of loss when coarse mesh was used had been underestimated. Their seminal work spawned over the next 30 years a plethora of studies on the influence of screen-mesh size on recovery (see James [1997] for a relatively complete listing of references as of a decade ago). Such studies continue to this day (e.g., Nagaoka 2005b; Partlow 2006), sometimes with much more statistical sophistication than that found in the original studies (e.g., Cannon 1999). Although the lessons learned have been significant ones, many of them were learned with Thomas's (1969) seminal effort. For that reason, various analysts have subsequently used his data to substantiate arguments concerning the influence of screen-mesh size on recovery (e.g., Casteel 1972; Grayson 1984).

Thomas (1969) used zooarchaeological data from three sites; this demonstrated that recovery was not simply a function of the particular sample (geographic and geological location) chosen. As each site was excavated, sediment was passed through a series of nested screens with increasingly finer mesh. The first screen was 1/4-inch (6.4 mm) mesh, the second was 1/8-inch (3.2 mm) mesh, and the final screen was 1/16-inch (1.6 mm) mesh. All faunal remains in each screen were retrieved and recorded as to screen mesh in which they were found. After all remains were identified, Thomas categorized the remains as to average adult live weight of an individual of the taxon represented. He distinguished five size classes: Class I: live weight < 100 g (e.g., mice); Class II: live weight 100 to 700 g (e.g., squirrels); Class III: live weight 700 g to 5 kg (e.g., rabbits); Class IV: live weight 5 to 25 kg (mid-size mammals); and Class V: live weight > 25 kg (e.g., deer). Thomas retained distinctions between site-specific samples, and also those between each vertical analytical level within each site. Such distinctions are irrelevant to studies of the loss of faunal remains, so we can ignore them and lump all data into categories defined by screen-mesh size and body size (Table 4.4).

Table 4.4. *Mammalian NISP per screen-mesh size class and body-size class for three sites. Percentages are calculated for each body-size class. Data from Thomas (1969)*

Body-size class	1/4 inch (%)	1/8 inch (%)	1/16 inch (%)	Total
I (< 100 gm)	141 (5)	910 (31)	1,930 (64)	2,981
II (100–700 gm)	626 (14)	1,478 (33)	2,450 (53)	4,554
III (0.7–5 kg)	1,069 (29)	1,358 (37)	1,275 (34)	3,702
IV (5–25 kg)	85 (96)	4 (4)	0	89
V (> 25 kg)	1,308 (100)	1 (0.1)	0	1,309
Total	3,229	3,751	5,655	12,635

Understating the issue, Grayson (1984:170) noted that there are “a number of ways in which [Thomas’s] recovery data can be analyzed, but no matter how the analysis proceeds, the effects of screen-mesh size on recovery are dramatic.” Although it doubtless is untrue, assume, for instance, that 100 percent of all faunal remains were recovered by the 1/16-inch screen mesh. We can then determine the cumulative percentage of NISP of each body-size class of mammal that was recovered across the increasingly finer screen-mesh size classes. These cumulative percentages are all plotted in Figure 4.7. That figure shows that the larger the body-size class is, the more of a taxon’s remains are recovered in coarse mesh screens, and the smaller the body-size class, the more of a taxon’s remains are recovered in fine-mesh screens. Thomas’s data empirically demonstrated what had long been suspected prior to his study – remains of small organisms are lost through coarse-mesh screens – and they demonstrate it with remarkable clarity. They demonstrate it on at least an ordinal scale because screen-mesh size classes and body-size classes are treated in Figure 4.7 as ordinal-scale variables. The one thing that we do not know from Thomas’s data is the nature of what is lost through the 1/16-inch mesh screens. But even without such information, Thomas’s data should prompt us to worry about taxonomic abundance data even if we use a fine-mesh hardware cloth, such as 1/8-inch or 1/16-inch mesh. Small taxa will be underrepresented relative to large taxa even when fine-mesh sieves are used. Deciding how thorough to be in recovery efforts (finer mesh will result in greater thoroughness) is a tactical decision that will depend on the research question asked and its attendant target variables.

Even though numerous empirical studies indicate that the coarser the screen mesh, the more small specimens pass through the sieve and are not recovered, occasionally this does not seem to hold true (e.g., Vale and Gargett 2002). The potential reasons for this are several (Gobalet 2005; Zohar and Belmaker 2005), but the most likely ones

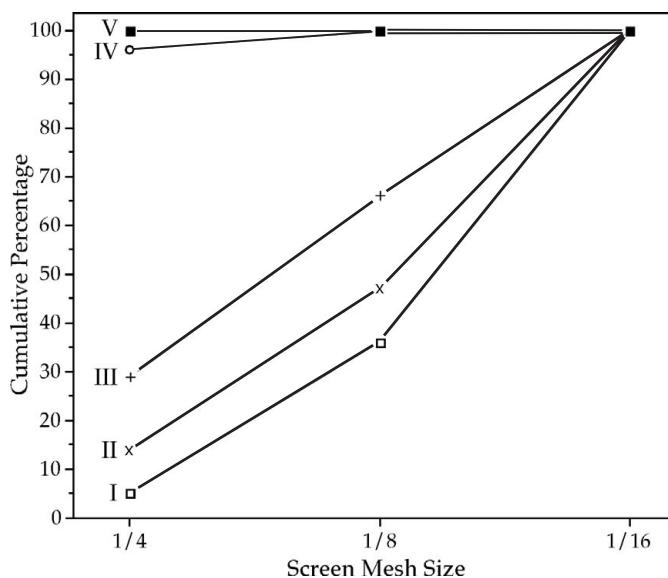


FIGURE 4.7. Cumulative percentage recovery of remains of different size classes (Roman numerals) of mammals. The critical but empirically unvalidated assumption is that all remains will be caught in the 1/16-inch mesh screen. Data from Table 4.4 (originally from Thomas 1969).

are taphonomic (Gargett and Vale 2005). If small remains are taxonomically unidentifiable because they are anatomically incomplete due to fragmentation, corrosion, or some other taphonomic process, then it is possible that the use of small sieves will not increase the value of NTAXA (e.g., Cooper et al. 2006). This is an empirical matter; every collection is unique and subject to investigation as to whether or not fine mesh makes a difference.

#### *To Correct or Not to Correct for Differential Loss*

If one passes site sediments through coarse-mesh hardware cloth, it is likely that small bones and small teeth will, like the sedimentary particles themselves, pass through the screen and thus not be recovered. The coarser the mesh of the hardware cloth – the larger the openings – the more remains of, first, small animals, and then progressively larger animals, as coarseness increases, will be lost because they are able to pass through the hardware cloth. The total magnitude of such loss will depend on the population of remains of small animals present in the screened sediments (Clason and Prummel 1977). The choice of sieve mesh size should depend on the

research questions one is asking because using finer mesh means it will take longer (and cost more) to complete an excavation – there will be more material caught in the screen that must be looked over and from which faunal remains must be removed.

One way to avoid the total cost of using fine-mesh sieves throughout an excavation is to take bulk samples every so often (how often is a matter of choice within the sampling design used) and to pass those bulk samples through one or more finer meshed sieves to determine what and how much is being lost. Some analysts have argued that if the rate of loss can be determined, then what has been recovered can be mathematically adjusted to account for what has been lost (e.g., James 1997; Thomas 1969; Ziegler 1965, 1973). Because differential recovery is often a troublesome concern, it is worthwhile to review one way to correct for differential loss.

Thomas (1969) suggested that the analyst determine a correction factor to analytically compensate for differential recovery of small remains. This might involve first using a formula like this:

$$\text{Percentage of NISP lost} = 100 \left( \frac{\sum \text{NISP from fine-mesh or bulk samples}}{\sum \text{NISP from fine-mesh or bulk samples} + \sum \text{NISP from coarse mesh or standard recovery method}} \right).$$

Once the percentage lost is known, the inverse of the fraction lost (represented by the percentage lost) can be multiplied by what has been recovered to estimate what would have been recovered if there had been no loss. Alternatively, Thomas (1969) suggests simply calculating the recovery ratio using the formula:

$$\text{Recovery ratio} = \frac{\sum \text{NISP for all recovery methods}}{\sum \text{NISP for recovery method of interest}}.$$

This formula is used for each size class of taxa. Thus, using the data in Table 4.4 for illustrative purposes, the recovery ratios per size class are: I: 21.14 (2981/141); II: 7.27 (4554/626); III: 3.46 (3702/1069); IV: 1.05 (89/85); and V: 1.00 (1309/1308). This means that if one wanted to correct for differential recovery that resulted from use of different screen mesh sizes at these sites, then the NISP of size class I remains should be multiplied by 21.14, the NISP of size class II remains should be multiplied by 7.27, size class III by 3.46, size class IV by 1.05, and size class V by 1.

There is a critical assumption that must be granted if a correction protocol such as that described by Thomas is to be used. The assumption is that the rate of loss determined from the subsample is representative of the entire sample. The weakness

of the assumption is that the recovery rate will likely vary from recovery context to recovery context because faunal remains tend to not be randomly distributed throughout a site or throughout a stratum. Loss will not be stable but in fact will likely vary not only from site to site and from stratum to stratum, but also from horizontal context to horizontal context within a site or stratum. Few researchers have explored this potentiality of a nonhomogeneous distribution of faunal remains with real data. Thomas (1969) used statistical procedures to determine that there seemed to be minimal vertical variation in the distributions of faunal remains, and so had an empirical warrant to apply his correction factor across entire site collections.

Not all sites have homogeneous distributions of faunal remains, and thus it is ill advised to calculate a correction factor based on one excavation unit (whether horizontally distinct, vertically distinct, or both) and to then apply that correction to another unit to obtain, say, a site-wide value (e.g., Cannon 1999; Lyman 1992a; Shaffer and Baker 1999). Occasionally paleozoologists have noted the proportion of a deposit that has been excavated, and then estimated frequencies of taxa in the entire site or deposit (e.g., Lorrain 1968). Again, such an estimation procedure assumes that the density of NISP per unit of area or unit of volume observed applies to the entire site or deposit under study. As data presented by Cannon (1999) demonstrate, such an assumption should be empirically validated, else estimates of total site content will be in error.

### *Summary*

Thus far several issues with respect to generating collections of faunal remains have been touched on. The focus has been to describe how one might determine if a collection is representative of a target variable by determining if one has sampled to redundancy or not, to illustrate how a particular recovery technique might influence what is collected (hand picking and screen mesh size), and to argue that despite being able to calculate a recovery rate in a mathematically elegant fashion, to utilize that rate as a correction factor is unwise given the requisite assumption that faunal remains are homogeneously distributed over the sampled deposit(s). For the sake of simplicity, throughout the chapter the focus has been on samples from which one seeks to measure taxonomic richness, or NTAXA. But the arguments hold with equal force for taxonomic abundances and other quantitative measures of the taxonomic composition of a collection, as demonstrated in Chapter 5.

The arguments made here also hold for nontaxonomic quantitative measures. For example, if the remains of taxa comprised of small individuals are lost more often

than the remains of taxa comprised of large individuals (Shaffer 1992), then it stands to reason that such intertaxonomic variation in recovery likely also applies intrataxonomically. In particular, small skeletal elements of a taxon will be lost more often than large skeletal elements (e.g., Nagaoka 2005b). Similarly, small fragments will be lost more often than large fragments (Cannon 1999). In general, small specimens will be lost more often than large specimens, regardless of the taxonomy or anatomical completeness of those specimens. The general lessons from such observations are two.

The first lesson rests on the fact that a relationship between sample size and the variable of interest may exist, so paleozoologists should search for such relationships (e.g., Koch 1987). If a relationship is found, then although the sample might in fact be representative of the variable of interest, the observed value of that variable might be result of sample size (Leonard 1997). Until such possible sample-size effects are controlled for analytically, or the relationship is found to be merely a correlation and not causal, it is ill-advised to interpret the variable in terms of some ecological or anthropogenic factor. The second lesson is that virtually any conceivable quantitative variable that *can* correlate with NISP will display values that are also potentially a function of sample size. Finding correlations between target variables and sample sizes does not preclude analysis and interpretation, but such findings suggest that cautious interpretation is warranted if the sample-size effects cannot be analytically controlled or eliminated. This brings up the important topic of how we might detect sample-size effects and how we might control for them.

#### THE SPECIES–AREA RELATIONSHIP

Botanists recognized in the early twentieth century that the larger the area they sampled the more species of plant they identified (Leonard 1989). Initially the relationship was thought to be linear – that as the area sampled increased, the number of species would increase at a constant rate. Within a decade or two it was empirically demonstrated that the relationship was semilogarithmic when large areas were considered. The number of species identified increased as the logarithm of the area increased. By the late 1930s, the relationship between amount of area sampled and number of plant species identified was being graphed as shown in Figure 4.8 (after Cain 1938). Within a few years, a graph of like form was generated for animal taxa but instead of the area sampled the independent variable was the total number of individual animals tallied (Fisher et al. 1943). The relationship between area examined and the number of taxa identified (NTAXA), and that between number of individuals tallied and NTAXA are the same because the more area examined the more individuals (whether

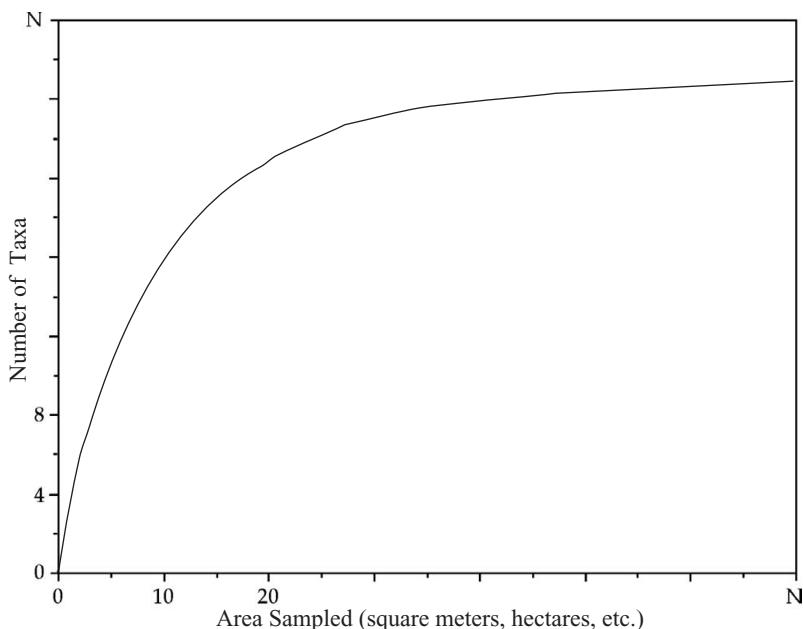


FIGURE 4.8. Model of the relationship between area sampled (or sampling intensity) and number of taxa identified.

plants or animals) are encountered. In ecology, graphs with the form of Figure 4.8 are sometimes referred to as accumulation curves. They are more often referred to as “species-area curves” because of the seminal discovery of the relationship between these two variables.

Given the nature of the relationship between the two variables, ecologists in the middle of the twentieth century became concerned with determination of how much area to sample, or how many individuals to tally, to ensure that their samples were representative of the target variable (often a habitat or biological community of some scale). One solution was to hold the area sampled constant at some minimum size thought to be adequate. Another is an analytical procedure termed “rarefaction” (Sanders 1968). *Rarefaction* involves determination of the number of species expected if all samples were the same size (if all samples included the same number of individuals). Richness or NTAXA for a fraction of a collection can be estimated by drawing a (random) subsample (equal to the fraction) of a sample (equal to the collection) of the population of interest. As Tipper (1979) states in his terse history (with pertinent references as of the late 1970s), the method is termed “rarefaction” because it involves reducing or rarefying a sample to a smaller size. Figure 4.9 illustrates the basic procedure and outcome of rarefaction in two ways.

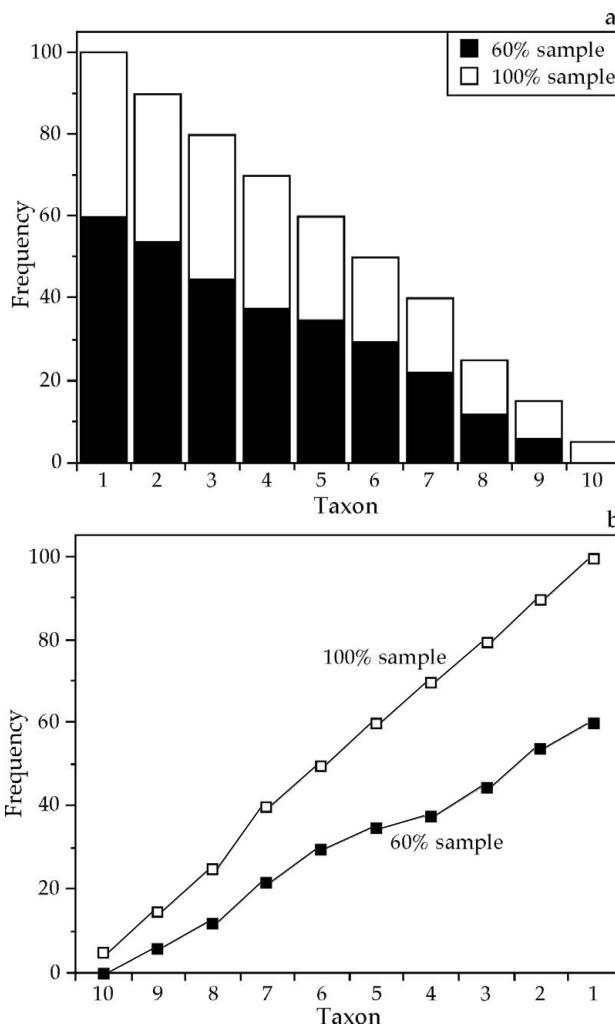


FIGURE 4.9. Two models of the results of rarefaction. (a) histogram with high white bars of 100 percent sample and black lower bars of rarefied (60 percent) sample; (b) rarefaction curve (compare with Figure 4.8) showing 100 percent sample and corresponding 60 percent rarefied sample.

Ecologists have been grappling with rarefaction for decades – its various forms and how to make results more valid (e.g., Colwell and Coddington 1994; Colwell 2004; Colwell et al. 2004; Gotelli and Colwell 2001; Scheiner 2003; Schoereder et al. 2004; Smith et al. 1985; Wolda 1981). Zooarchaeologists have been aware of the basic rarefaction procedure for more than 20 years (Styles 1981), although few individuals have used it (see Lyman and Ames 2007 for references). Paleobiologists are also aware of the method, and they have devoted considerable effort to developing and

perfecting it (e.g., Alroy 2000; Barnosky et al. 2005; Bush et al. 2004; Miller and Foote 1996). Early efforts to develop standard species area curves for paleozoology (Koch 1987) have not been pursued, probably because general patterns are too general to be of predictive value.

Given that the basic rarefaction procedure involves reducing a sample to a smaller size, it is not surprising that as the statistical sophistication of scientists increased and access to electronic computing power increased in the 1970s, programs were written explicitly to perform rarefaction analysis. The best known of these among zooarchaeologists is one designed by Kintigh (1984). This procedure sums all available samples in order to model taxonomic abundances in the population, and then draws random samples of various sizes from that modeled population. Richness is determined multiple times for each sample size, and a mean richness and confidence levels thereof are calculated for each sample size. Finally, the procedure generates not only a best-fit line (mean) through the sample data sets but also confidence intervals for the line in graphic form. This rarefaction program has been used by zooarchaeologists to compare faunas of different sizes (e.g., McCartney and Glass 1990). The resulting model approximates the effects of varying sample size on richness and is designed to test the null hypothesis that all samples (of whatever size) were derived from the same population, and thus to identify samples that are not members of the population but are instead (statistical) outliers. An *outlier* is a sample that seems not to have been drawn from the same population as all others because it falls far above or below the richness expected given its size; an outlier is a sample that, probabilistically, could not have been drawn from the modeled population.

Several seldom acknowledged assumptions and problems attend rarefaction. Early on, Grayson (1984:152) noted that the rarefaction method in general as originally developed by Sanders (1968) and later perfected by Tipper (1979) used quantitative units that were statistically independent of one another; it used individual animals. No similar quantitative unit is available for paleozoology. One might use MNI, but these values are dependent on aggregation; one might use NISP, but these values are likely interdependent to some unknown degree.

Rhode (1988) noted that if one uses Kintigh's (1984) procedure (and null hypothesis) then one is assuming that a great deal is already known about the population being investigated. In particular, such use assumes that the samples used to generate the rarefaction curve are, when summed, representative of taxonomic richness as manifest in the population of interest and, more importantly, that their sum is also representative of the distribution of individuals across taxa (known as taxonomic evenness). Using the sum of all samples to generate a rarefaction curve such as in Figure 4.9b may result in the inclusion of samples that are not members of the (target)

population; if the samples derive from different populations, their sum will represent a sample of organisms derived from those multiple populations. The statistical effect of including all samples is to produce expected richness values for various sample sizes that have been influenced by one or more samples that may not actually be part of the same (target) population (the same holds for taxonomic evenness). Differences between a nonmember sample and the model generated from all samples including the nonmember would be muted to some unknown degree (see also Byrd 1997). As Rhode (1988:711–712) astutely observes, if a particular sample used to model the target population seems to differ significantly from that modeled population, how can “the choice of that population as the comparative baseline be justified?”

As Kintigh (1984) originally noted, the key step in his rarefaction procedure involves the definition of the population; in particular, which samples are to be included when summing samples to create the population model? Producing an answer to this question is where the assumption that we already know much about the population we are studying comes into play. Analytical means of evaluating whether samples of different sizes might have been derived from the same underlying population are discussed later in this chapter.

On the one hand, Buzas and Hayek (2005) recently defined “within-community sampling” as drawing  $>1$  sample from a population with a particular frequency distribution (set of taxonomic abundances) or constant value of a variable of interest. “Between-community sampling,” on the other hand, involves drawing the  $>1$  samples from populations with different frequency distributions or the same distribution with different values of a variable of interest. The distinction could be used as a basis for lumping two samples (they are statistically indistinguishable with respect to the property of interest) or for not lumping two samples (they are statistically distinct).

The preceding returns us to the question of what constitutes the target variable? If it is NTAXA in a biological community, how is the community defined (see Chapter 2)? If it is the taxa exploited by human occupants of an archaeological site, the differences between the thanatocoenose, taphocoenose, and identified assemblage must be kept in mind (Figure 2.1). This volume is not the place to explore these issues. Rather, it is relevant to illustrate how analysts have studied and analytically used the generic species-area relationship. To do that in the following, it is assumed that the target variable is NTAXA within the identified assemblage. This simplification allows us to focus on the species-area relationship and methods of investigating it, although it is important to note that the relationship may well be found to exist between any measure of sampling effort or sample size and any target variable (richness, evenness, heterogeneity).

*Species–Area Curves Are Not All the Same*

In preceding pages, techniques to explore relationships manifest by species–area curves have been mentioned (e.g., Figures 4.1–4.4 and 4.8–4.9, and associated discussion). At this juncture it must be made clear that species–area curves do not all express the same relationships or have the same implications with respect to the relationship between sample size and NTAXA. This is so because they are constructed differently, and they are constructed differently because they have different analytical purposes and address different analytical questions. To demonstrate this, in the following the data in Table 4.2 are used to construct three different kinds of what are generically known as species–area curves. (A portion of this section is derived from Lyman and Ames [2007].)

One kind of species–area curve is shown in Figure 4.2. In this curve samples increase in size by being added together and thus are statistically interdependent. This kind of species–area curve is a sampling to redundancy curve. The particular curve in Figure 4.2 has leveled off, suggesting that all of the information in the last couple samples (identities of the mammalian genera present) is redundant with information provided by earlier (smaller) samples. If the curve had not leveled off, such as is the case in Figure 4.3, then new samples are still adding new information so there is no empirical basis to argue that we have sampled to redundancy. The sampling to redundancy curve can be plotted manually by simply connecting points, or it can be drawn statistically (Lepofsky and Lertzman 2005). A sampling to redundancy curve has a very narrow analytical purpose – to determine if increases in sample size (accomplished by summing samples) influence the target variable; its utility is that it provides an empirical indication of sample adequacy in the form of a static value for the target variable across samples of varied sizes that comprise one total collection. Constructing a species–area curve of the sampling to redundancy kind is straightforward, but remember that the order of sample addition will influence the ultimate sample size at which the curve levels off (see the discussion of Figures 4.2 and 4.3).

Many species–area curves have been constructed in one of two ways distinctly different from how a sampling to redundancy curve is built. They are distinct because they have different analytical purposes. Some of those other curves were constructed to compare statistically independent samples of different sizes (e.g., McCartney and Glass 1990); some were constructed from statistically independent samples derived from one population in order to predict representative statistically independent sample sizes drawn from other populations (e.g., Zohar and Belmaker 2005); some were used to determine or compare rates of increase in richness (slope of the curve)

(e.g., Grayson 1998); some were constructed by rarifying samples (reducing their sizes probabilistically) (e.g., Styles 1981). How were the other curves constructed?

One way that species-area curves are constructed involves generating bivariate plots of statistically independent samples, and then statistically fitting a curve to the plot to determine if sample size may be influencing the target variable across the different samples. The example in Figure 4.4 uses the six annual samples from the Meier site described in Table 4.2. The best-fit regression line defined by the point scatter is included. The correlation and the regression line are statistically significant ( $p < 0.01$ ) and suggest that NTAXA per statistically independent annual sample is a function of sample size measured as NISP. If each point represented a sample from a different stratum or different site, Figure 4.4 would suggest those samples were strongly influenced by sample size, and thus NTAXA values for those samples should not be compared.

Figure 4.4 does not allow us to surmise if our total sample from the Meier site is representative of taxonomic richness (compare with Figure 4.2); the kind of curve in Figure 4.4 has a different analytical purpose and utility. The protocol of building a species-area curve exemplified in Figure 4.4 is sometimes referred to as the “regression approach” (Leonard 1997). The name reflects the statistical analysis performed. Regression analysis ascertains the strength of the relationship between samples of different sizes (in Figure 4.4, NISP values) and a target variable (in Figure 4.4, NTAXA). The strength of the relationship is reflected by the magnitude and statistical significance of the correlation coefficient. If there is a significant correlation between sample size and the target variable, then the magnitudes of the target variable could be a result of sample sizes rather than a property of interest. With respect to Figure 4.4, taxonomic richness varies according to sample size. Therefore, if these samples had come from different strata or sites, we would not want to conclude something like the sample from the 1991 site/stratum is taxonomically richer than the 1990 site/stratum, so the people who deposited the remains in the 1991 site/stratum had greater diet breadth than those who deposited the 1990 site/stratum materials. Remembering that correlations do not necessarily imply a causal relationship between two variables, our inference regarding diet breadth might be correct, but it might not. The regression approach is merely a way to detect those instances when caution is advisable.

If the regression approach prompts the conclusion that sample-size effects may be present in a set of samples, the analyst has options. The samples can be pooled and a rarefaction analysis performed, if one is willing to make the necessary assumptions. Alternatively, slopes of lines describing the relationship between sample size and the target variable may vary across different sets of samples (see Chapter 5). Comparisons of slopes may reveal a property of the compared sets of samples not otherwise

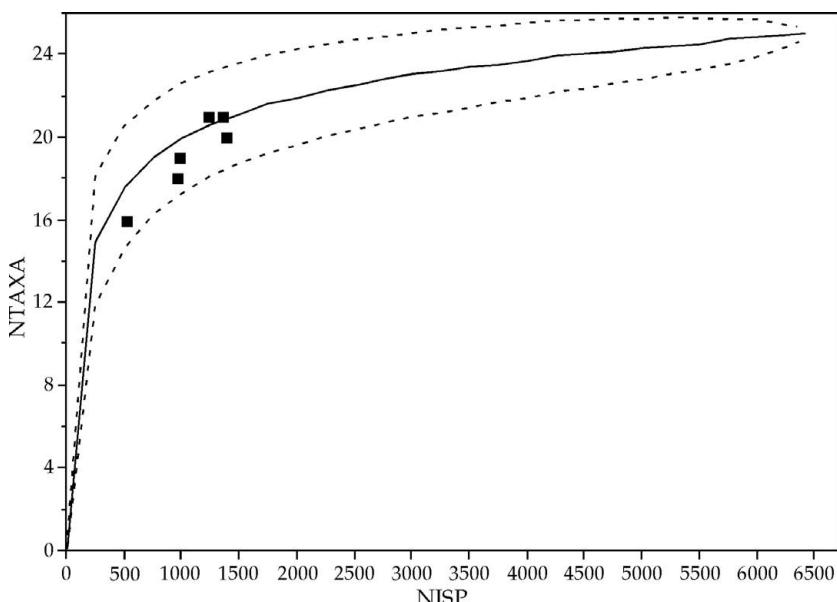


FIGURE 4.10. Rarefaction curve (solid line) and 95 percent confidence intervals (dotted lines) of richness of mammalian genera based on six annual samples from the Meier site (black squares). Data from Table 4.2; curves determined using Holland's (2005) *Analytical Rarefaction*.

detectable that is free of sample-size effects. A third possibility is to identify statistical outliers, or samples that fall significant distances (usually  $\geq 2$  standard deviations) from the regression line (Grayson 1984). Ascertaining why samples fall far from the regression line may reveal a unique property of those unusual assemblages not otherwise perceived and that is free of sample-size effects. Study of slopes and of outliers avoids one weakness of the regression approach. Small samples may in fact be 100 percent samples or populations (Rhode 1988), and thus the sample-size effect is an artifact of the size of the populations from which the samples derive.

The third way that species-area curves are constructed involves *rarefaction* (e.g., Sanders 1968; Tipper 1979). Rarefaction has been used by zooarchaeologists for some time (e.g., Byrd 1997; McCartney and Glass 1990; Styles 1981). There are several ways to construct rarefaction curves, but describing them is beyond my scope here. It suffices to say that one can use statistically independent samples or statistically interdependent (summed) samples (or sample without replacement, or sample with replacement) to estimate NTAXA were a sample of a particular size. A rarefaction curve constructed using the six annual samples from the Meier site is shown in Figure 4.10. To generate this curve, Holland's (2005) *Analytical Rarefaction* software

was used. If the six samples from Meier were independent of one another and from different strata or sites, the rarefaction curve would allow comparison of NTAXA across assemblages of different size without fear of sample size differences driving the results. As noted earlier, the rarefaction procedure assumes the included samples all derive from the same population, and it also assumes that specimens used to provide (NISP) values for drawing the curve are independent of one another. In Figure 4.10, we know the samples all derive from the same population (the Meier site), and thus we also know the specimens are to some degree interdependent.

The three kinds of species-area curves shown in Figures 4.2, 4.4, and 4.10 are not very similar in general appearance despite the similarities in the variables used to build them. They are not very similar because each curve is meant to address a distinct analytical question, so each has been built in a unique, distinctive way. The sampling to redundancy approach (Figure 4.2) determines if one total collection represents the value of the target variable. Regression analysis (Figure 4.4) allows detection of possible sample size effects on the target variable among independent samples of different size. Rarefaction (Figure 4.10) allows two or more samples of different sizes to be compared as if they were the same size by reducing the larger samples to a common small size.

## NESTEDNESS

There is an analytical means of evaluating whether samples of different sizes might have been derived from the same underlying population. The analytical technique was developed by biogeographers studying insular faunas such as those on archipelagos or island chains (see Brown and Lomolino [1998] for details). They reasoned that the faunas on land-bridge islands (those once connected to the mainland when sea levels were low) likely originated on the mainland, and given the species-area relationship, islands – which have varied but relatively small land areas – would have subsets of the taxa found on the mainland – which have large land areas relative to islands. Further, small islands would have smaller subsets of taxa – have lower NTAXA values – than would large islands. Islands can be oceanographic, or they can be habitat islands surrounded not by water but habitats unfavorable to the taxa located in the insular habitat patch. The pattern of organismal distribution – presence/absence of taxa across the islands – is referred to as the “nested subset pattern” (Patterson and Atmar 1986; see also Cutler 1994; Patterson 1987; Wright et al. 1998).

The concept of a nested subset pattern is straightforward. Figure 4.11 shows both a perfectly nested set of faunas, and a poorly nested set of faunas, in two graphic

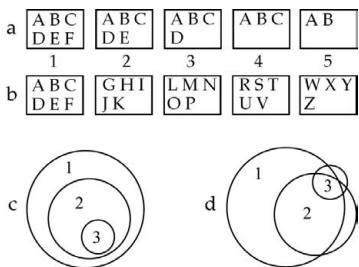


FIGURE 4.11. Examples of perfectly nested faunas and poorly nested faunas. (a) perfectly nested set of faunas, each capital letter represents a unique species; (b) poorly nested set of faunas, each capital letter represents a unique species; (c) Venn diagram of three perfectly nested faunas in which the larger the circle, the greater the number of taxa; (d) Venn diagram of three imperfectly nested faunas in which the larger the circle, the greater the number of taxa. (a) after Cutler (1994); (c) and (d) after Patterson (1987).

forms. Table 4.5 shows a perfectly nested set of faunas and a poorly nested set of faunas in tabular form. In the perfectly nested sets, taxa absent from one fauna are also absent from all smaller faunas, and taxa present in a fauna are also present in all larger faunas. In poorly or weakly nested faunas, some taxa may occur unexpectedly in small faunas and large faunas but not in midsized faunas, and other taxa may not occur in large faunas but occur in midsized or small ones. The unexpected occurrences are “outliers” whereas the unexpected absences are “holes” in the nested pattern (Cutler 1991).

The extremes of nestedness are easy to tell apart (Figure 4.11, Table 4.5). What about intermediate cases? Can we determine if one set of faunas is more nested than another? Biogeographers have developed quantitative ways to measure exactly how nested a set of faunas is, and thus one can compare the nestedness of multiple sets of faunas (e.g., Cutler 1991). Atmar and Patterson (1993) refer to their algorithm for measuring the degree of nestedness as a means to measure an archipelago’s “heat of disorder” or “temperature.” The algorithm measures the *degree* of nestedness on a scale of zero to 100 degrees; faunas that are perfectly nested have a temperature of  $0^\circ$  whereas faunas that display no nestedness whatsoever are  $100^\circ$ . (The 100 degrees are an arbitrary interval-scale measure of amount of nestedness.) The value of the nestedness concept is great because, theoretically, nestedness provides an indication of whether two or more faunas derive from the same population. In a way, the examination of nestedness is like rarefaction without rarefying; it compares samples rather than sum them and rarify the sum.

Atmar and Patterson’s (1993) thermometer of nestedness provides a measure of whether multiple faunal (island) samples derive from the same underlying

Table 4.5. Two sets of faunal samples showing (a) a perfectly nested set of faunas and (b) a poorly nested set of faunas. +, taxon present; -, taxon absent. (b) was generated with a table of random numbers

Assemblage	Taxon A	B	C	D	E	F	G	H	I	J
a. Nested										
I	+	+	+	+	+	+	+	+	+	+
II	+	+	+	+	+	+	+	+	+	-
III	+	+	+	+	+	+	+	+	-	-
IV	+	+	+	+	+	+	+	-	-	-
V	+	+	+	+	+	+	-	-	-	-
VI	+	+	+	+	+	-	-	-	-	-
VII	+	+	+	+	-	-	-	-	-	-
VIII	+	+	+	-	-	-	-	-	-	-
IX	+	+	-	-	-	-	-	-	-	-
X	+	-	-	-	-	-	-	-	-	-
b. Not nested										
I	-	-	-	+	-	-	+	-	-	+
II	-	+	+	+	-	+	+	+	-	-
III	-	-	+	-	-	-	+	+	+	+
IV	+	-	-	+	-	+	+	-	+	+
V	-	+	-	-	+	-	+	-	-	+
VI	+	-	+	-	+	-	-	-	-	+
VII	-	-	-	+	+	+	-	-	+	+
VIII	+	-	+	-	-	-	-	+	-	-
IX	+	-	+	+	-	-	+	-	-	-
X	+	+	-	+	+	+	+	-	+	-

(mainland) population. If the faunas are strongly nested, then it is probable that the samples derive from the same population, and one might perform a rarefaction analysis using those faunas lumped together (assuming quantitative units are independent). If faunas are weakly nested, then one could argue that either the samples are so small as to either not accurately reflect the heterogeneity of the population or the samples derive from different populations. How strong must nestedness be, or how weak? That is difficult to answer. But the point is that the nestedness thermometer provides a measure that constitutes information bearing on the answer. And a well-informed decision is likely to be better than one that is poorly informed.

The nestedness diagram of the 18 assemblages from eastern Washington State generated by Atmar and Patterson's (1993) thermometer is shown in Figure 4.12.

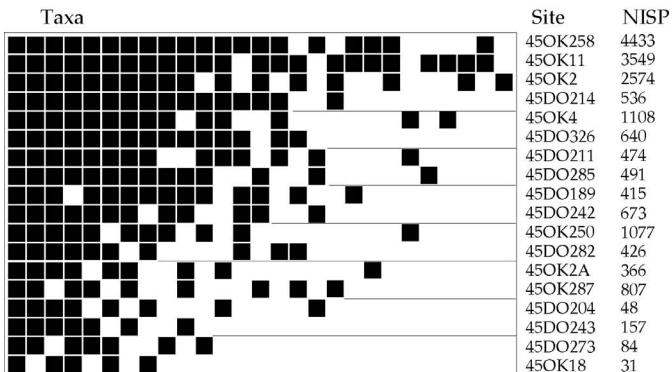


FIGURE 4.12. Nestedness diagram of eighteen assemblages of mammalian genera from eastern Washington State. Note that the NISP per assemblage and the rank order of the assemblages are strongly correlated (Spearman's  $\rho = 0.812$ ,  $p < 0.0001$ ).

That figure suggests there is some nestedness among the faunas. This set of faunas has a nested “temperature” of  $18.23^\circ$ , a value that suggests there is indeed some nestedness ( $0^\circ$  is perfectly nested), but the faunas are hardly perfectly nested. In conjunction with the facts that NISP and NTAXA values per assemblage for this set of assemblages are strongly correlated (Figure 4.4), and that the order of nested faunas produced by the nestedness thermometer is strongly correlated with NISP per assemblage ( $\rho = 0.812$ ,  $p < 0.0001$ ), it seems reasonable to conclude that all eighteen assemblages derive from the same population of mammals. The assemblages merely differ in size ( $= \sum \text{NISP}$ ), and that difference is the major variable that is creating taxonomic differences between them.

The value of the nestedness concept, however it is determined (and there are several ways to do so; compare Cutler [1991] with Atmar and Patterson [1993]), is great. If one grants the assumption that a small fauna should approximate a random sample of a large fauna, then when comparing two or more faunas of different sizes, if all faunas derive from the same population, they should be nested. The nestedness concept takes advantage of not only the relationship between sample size and NTAXA (say), but the taxonomic composition of the faunas. Rarefaction does as well, but it effectively begins with the assumption that the faunal samples are all from the same population. The nestedness concept and the techniques for measuring nestedness allow that assumption to be tested and evaluated empirically. Given that the concept has been discussed in the ecological literature for more than two decades, and given the near ubiquitous concern with sample size issues among paleozoologists, it is a bit surprising that nestedness has not been used by paleozoologists with some frequency. Indeed, I am aware of only one instance of a paleozoologist using it (Jones 2004).

## CONCLUSION

There is a particularly telling example in the recent literature that highlights the lack of interdisciplinary contact. Leonard (1987) mentioned the sampling to redundancy approach 20 years ago in the archaeology literature. That technique was mentioned, if not used very often, in the zooarchaeological literature several times since then (e.g., Lyman 1995a; Monks 2000; Reitz and Wing 1999). A recent analysis by paleontologists began with rarefaction to identify the general shape of a curve produced by samples of different sizes and NTAXA (Jamniczky et al. 2003). Those paleontologists discovered that the sampling to redundancy approach was a valuable tool for assessing sample representativeness. They seem unaware of any of the discussions in the zooarchaeological literature of their discovery. These paleontologists also do some impressive computer modeling to try to predict how many additional samples might be necessary, but this echoes Kintigh's (1984) method (which they do not reference), and they tend to caution against its use.

The paleontologists cited in the preceding paragraph made a significant contribution to paleontology and introduced an important analytical technique to that discipline. The point here is simple. Read some paleontology if you are a zooarchaeologist; if you are a paleontologist or paleobiologist, read some zooarchaeology. The cross-fertilization will be worthwhile.

The means by which a collection of faunal remains is generated – which sampling design is used, how large the sample is, how faunal remains are extracted from sediments – can and typically does influence what is recovered. Specifically, the size of the sample collected, and the frequencies of many of its attributes, are influenced by what (and how much) is collected. In the next three chapters, several different target variables are discussed. Throughout, analytical means of detecting and controlling for sample-size effects are described as quantitative measures of the target variables are sought. Given that the variables of interest are quantitative, analysts need to be aware of sample-size effects and to take every precaution to avoid allowing conclusions to be influenced by them. Means to detect such effects and a possible means to control for them have been described in this chapter. We will return to these analytical techniques often in Chapters 5, 6, and 7.

## Measuring the Taxonomic Structure and Composition (“Diversity”) of Faunas

One of the most common analytical procedures in paleozoology is to compare faunas from different time periods, from different geographic locales, or both (e.g., Barnosky et al. 2005 and references therein). Comparisons may be geared toward answering any number of questions. Does the taxonomic composition of the compared faunas differ (and why), and if so, by how much (and why)? Does the number of taxa represented (NTAXA) differ between faunas (and why), and if so, by how much (and why)? Do the abundances of taxa vary (and why)? Ignoring the “and why’s,” these and similar queries are what can be considered proximal questions. The why questions are the ultimate questions of interest; they constitute a reason(s) to identify and quantify the faunal remains in the first place. Was hominid or human dietary change over time the cause of the change in taxonomic composition, abundance, and so on? Did the environment (particularly the climate) change such that different ecologies prompted a change in the taxa present, the number of taxa present, or the abundances of various taxa? It is beyond the scope of this volume to consider these ultimate *why* questions other than as examples. The purpose of this chapter is to explore how quantitative faunal data can be analytically manipulated in order to produce answers to these kinds of proximal questions.

Once faunal remains have been identified as to the taxa they represent, they can be quantified or counted any number of ways, many of which are described in Chapters 2 and 3. As indicated in those earlier chapters, NISP tends to be the quantitative unit of choice for many analyses. NISP is used in this chapter to illustrate how taxonomic abundance data can be analytically manipulated in order to measure the taxonomic structure and composition of a collection of paleofaunal remains. MNI and biomass might also be used to calculate the indices described, but in many cases there are good reasons to not use them, as argued in earlier chapters.

Use of NISP throughout this chapter is meant to endorse it as the quantitative unit of choice in such efforts. This does not mean that NISP is without flaws that

might influence analytical results. Whether a set of NISP values for an assemblage suffers from interdependence should be ascertained prior to performing analyses like those described in this chapter. Methods to do this are described in Chapter 2. If the NISP values do not seem to be influenced by variation in interdependence, then use NISP values as ordinal scale values. If the NISP values are plagued by interdependence, then the data are best treated as nominal scale data. As we will see, even if interdependence does not seem to be a problem, there are other concerns with using NISP as an estimate of a property of a paleofauna.

The analytical gymnastics involving number of taxa, shared taxa, taxonomic abundances, and the like typically are implicitly aimed at the biocoenose (biological community) by paleobiologists, whereas zooarchaeologists may seek measures of the thanatocoenose (killed population) or the biocoenose depending on the research question. Recall that a biological community is a slippery entity empirically and conceptually. Allowing that a community can indeed be defined as, say, a naturally delineated habitat patch (if defined with even greater difficulty in the prehistoric record than in modern ecosystems), ecologists tend to recognize three levels of inclusiveness of biological diversity (Whittaker 1972, 1977). *Alpha diversity* is the diversity within a single local community; *beta diversity* is the change in diversity among or across several communities (recognizably distinct but adjacent habitats); and *gamma diversity* is the diversity evident in a set of communities such as is found across a large area (Loreau 2000) involving more than one kind of habitat. Paleozoologists do not always ignore these various sorts of diversity (e.g., Sepkoski 1988) when they compare diversity across geographic localities, temporal periods, or both, but sometimes they do (e.g., Osman and Whitlatch 1978). Of course sometimes they must assume (or analytically warrant the belief) that the samples they use are each derived from a single community and are not time and space averaged (e.g., Bush et al. 2004), though this is not always possible or necessary depending on the question they are asking (e.g., Jackson and Johnson 2001; Sepkoski 1997).

Given its central role in identifying the target variable, *diversity* is a concept in need of explicit definition. The title of this chapter is “Measuring the Taxonomic Structure and Composition (‘Diversity’) of Faunas.” This wording is meant to imply that *diversity* signifies the structure and composition of a fauna. By structure and composition is meant such variables as the particular taxa represented in a collection of faunal remains, the number of taxa represented regardless of which taxa are represented, the abundances of various taxa, and the like. In the ecological and biological literature *diversity* has come to mean any number of these variables (Magurran 1988; Spellerberg and Fedor 2003). In fact, some years ago the term “diversity” signified numerous concepts and variables within ecological research, and thus one

ecologist suggested that it be abandoned because it was too ambiguous (Hurlbert 1971).

The term “diversity” was not abandoned, but the lesson here is an important one. Be aware that the terms analyst A uses may have different meanings than those intended by analyst B. I follow precedent in zooarchaeology (e.g., Byrd 1997; McCartney and Glass 1990) and some ecological literature (e.g., Lande 1996), and use the term “diversity” to signify a family of variables used to describe the structure and composition of faunas and collections of faunal remains. The members of that family of diversity variables are introduced in the following section. In subsequent sections, quantitative indices for each variable are discussed. Although it is sometimes obvious that one fauna differs in one or more ways from another fauna, the indices have been designed to provide a quantitative measurement of similarities and differences. Many of these indices provide a continuous measure of similarity or difference, and this facilitates comparative analyses.

#### BASIC VARIABLES OF STRUCTURE AND COMPOSITION

The simplest variable that measures a property of a fauna is the number of identified taxa (NTAXA). NTAXA is often referred to in the ecological literature as number of species or species richness (Gaston 1996), but in fact the number of taxa in a fauna can be tallied at any taxonomic level so long as only one level is tallied. Mixing taxonomic levels and summing them to determine NTAXA would produce results that have unclear meaning, particularly when comparing the structure and composition of faunas using the variables introduced in this chapter. For example, recall that to be taxonomically identifiable, a bone must retain taxonomically diagnostic features and if that bone is fragmentary and thus anatomically incomplete, then it may not be identifiable to, say, the species level but only to the genus level. Thus, differences in richness tallied at multiple taxonomic levels may actually measure the degree of identifiability (and fragmentation) rather than taxonomic richness.

If taxa of different taxonomic levels are summed, then the same taxon may be counted twice. Consider, for instance, the fact that not all deer bones can be identified to species, but some can be so identified (Jacobson 2003, 2004). In western North America there are two species of deer – *Odocoileus virginianus* and *Odocoileus hemionus*. If in a collection of paleozoological remains some remains could be identified to one species, other remains could be identified to the other species, and still other remains could only be identified to genus but not species, then to tally all three

would be to count one (or perhaps both) species twice – once at the species level and once at the genus level (Grayson 1991a). By limiting NTAXA tallies to a single taxonomic level, the variable being measured is more self evident than were taxa of varied levels summed, and we have not risked counting the same phenomenon twice. NTAXA is a tally of the number of taxa identified; it is a nominal scale measure of taxonomic abundances (taxa are present or absent). Whether or not NTAXA can be a ratio scale measure of the number of taxa, or even an ordinal scale measure, is one of the topics of this chapter.

Stock's (1929) data for Rancho la Brea indicated an NTAXA of five mammalian orders (see Chapter 1). The sample of owl pellets mentioned in earlier chapters (Table 2.9) has a richness of mammalian genera of six (= NTAXA). The Meier site faunal remains have an NTAXA of twenty-six mammalian genera, and the complete assemblage from the Cathlapotle site has an NTAXA of twenty-five mammalian genera (Table 1.3). Over the years, ecologists and biologists have referred to this variable as taxonomic richness (Gaston 1996; Odum 1971; Palmer 1990), taxonomic variety (Odum 1971), taxonomic density (Pianka 1978), and taxonomic diversity (Brown and Gibson 1983; Colwell et al. 2004; Ricklefs 1979; Spellerberg and Fedor 2003). The term "taxonomic richness" here signifies NTAXA. *Diversity* is used as a generic term for several variables, including richness. Later in this chapter taxonomic density is mentioned; just as is implied by the term "density," this measure is a rate or ratio (e.g., NTAXA/ $\sum$ NISP).

Paleofaunal samples (or biological communities) may have similar NTAXA values. Even if they do not have similar NTAXA values, the structure and composition of the faunas could vary in terms of taxonomic abundances. The several taxa of one fauna may all have approximately equal abundances, such as in the case of a fauna with ten taxa and each constitutes 10 percent of the total individuals (or biomass). The taxa of another fauna may have rather unequal abundances of each of ten taxa, such as three taxa each representing about 1 percent of the total individuals, another three taxa each representing about 5 percent of the total individuals, another three each representing about 10 percent of the total individuals, and the tenth taxon representing the remaining 50–52 percent of the individuals. In such cases the first fauna described is said to be taxonomically even whereas the second fauna described is taxonomically uneven. *Taxonomic evenness*, sometimes referred to as taxonomic *equitability* (Art 1993; Magurran 1988), is a measure of how individuals are distributed across categories, in this case, taxa (Smith and Wilson 1996). Faunas are taxonomically *even* if each taxon has the same number of individuals (or whatever variable is used to measure abundances of taxa) as every other taxon, regardless of richness. Faunas are taxonomically *uneven* if each taxon has a different number of individuals than every

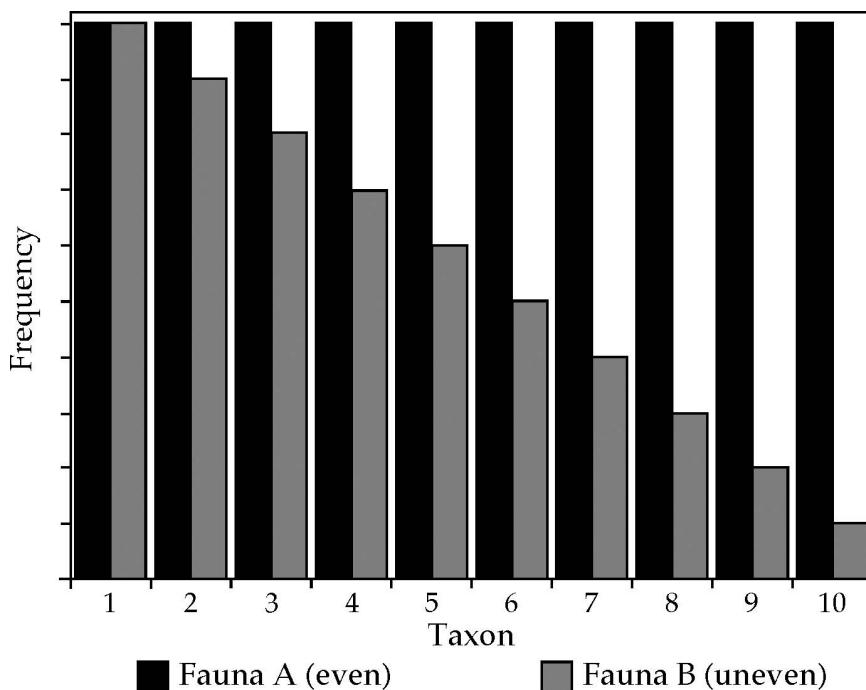


FIGURE 5.1. Two fictional faunas with identical taxonomic richness (NTAXA) values but different taxonomic evenness.

other taxon, regardless of richness. Indices for measuring evenness will be introduced in this chapter.

Figure 5.1 shows two faunas with the same taxonomic richness: NTAXA = 10. But those two faunas differ considerably in terms of how individuals are distributed across taxa, so evenness varies regardless of richness. Can both richness and evenness be measured simultaneously? Of course. Characterizations of the structure and composition of a fauna that measure richness and evenness simultaneously are sometimes referred to as measures of taxonomic diversity (Magurran 1988; Odum 1971; Pianka 1978; Ricklefs 1979; Spellerberg and Fedor 2003) or heterogeneity (references in Peet 1974). Recall that the term “taxonomic diversity” has had (and continues to have) many meanings. As Peet (1974:285) observed, “diversity has always been defined by the indices used to measure it.” The term *heterogeneity* is used in the following to signify simultaneous measurement of evenness and richness to avoid confusion with how the term “diversity” is used in this chapter.

A description of heterogeneity that may help conceptualize what the notion means underscores that the term signifies the simultaneous measurement of both evenness and richness. Pianka (1978:287) used the term diversity the way that heterogeneity

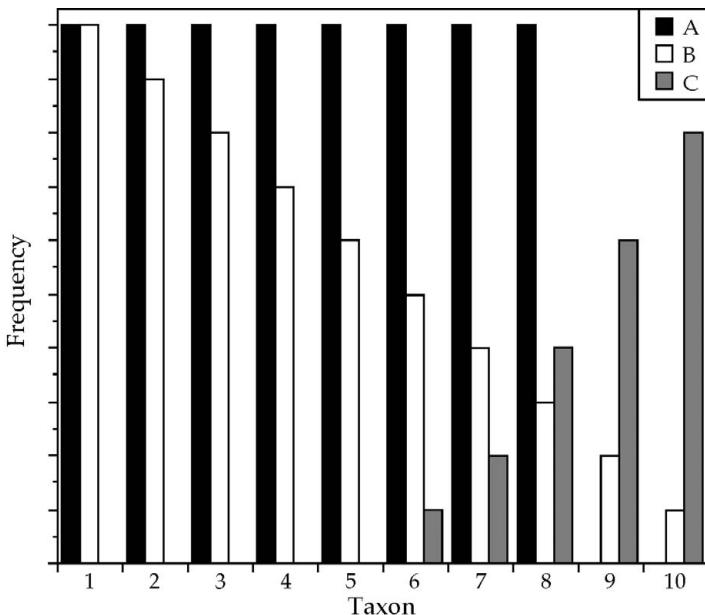


FIGURE 5.2. Three fictional faunas (A, B, C) with varying richness values and varying evenness values.

is used here, and stated that taxonomic heterogeneity “is high when it is difficult to predict the species of a randomly chosen individual organism and low when an accurate prediction can be made.” Thus, as NTAXA increases, the predictability of the taxonomic identity of any single randomly chosen individual decreases; it is easier to predict which taxon is represented if only two taxa are possible (you have a one out of two chance, or a 50 percent chance) than it is if ten taxa are possible (you have only a one out of ten chance, or a 10 percent chance). And, if NTAXA is two, but species A is represented by ninety-nine individuals and species B is represented by only one individual, the predictability of the taxonomic identity of any single randomly chosen individual is high. There is a 99 percent chance a randomly chosen individual will be taxon A but only a 1 percent chance it will be taxon B. Thus as richness increases, as evenness increases, or both, heterogeneity increases (because the predictability of the taxonomic identity of a randomly chosen individual decreases). Figure 5.2 illustrates three fictional faunas with varying richness values and varying evenness values. Think about randomly drawing a single individual from any one of these faunas and how often you could correctly predict the taxonomic identity of that individual.

Taxonomic richness (NTAXA) is directly correlated with heterogeneity (both either increase, or decrease, together), and taxonomic evenness is also directly correlated with heterogeneity (both either increase, or decrease, together). In this book the

term taxonomic heterogeneity means just what Pianka (1978) and others (Peet 1974; Spellerberg and Fedor 2003) mean when they use the term – a combined measure of NTAXA and taxonomic evenness. In later sections of this chapter, several quantitative measures of heterogeneity are introduced that utilize taxonomic abundance data. But before those indices are described and exemplified, indices of richness need to be discussed. As well, indices that measure the degree of similarity of two faunas need to be described. We start with simple indices of structure and composition, and move to more (mathematically and conceptually) complex indices.

#### INDICES OF STRUCTURE AND SIMILARITY

Two faunas can be compared in terms of several different variables a particular value of which each fauna displays. These variables are (1) NTAXA or taxonomic richness (regardless of which taxa are represented), (2) taxonomic composition (the particular taxa represented), (3) taxonomic heterogeneity, and (4) taxonomic evenness. They are discussed in the order listed because in that order, complexity (both mathematical and information content) increases and indices introduced later in this section tend to rest on indices introduced early.

Taxonomic composition was not mentioned in preceding paragraphs because richness, heterogeneity, and evenness are, in a sense, taxon free; their values in any given instance will vary regardless of the taxa present. As paleobiologist Thomas Olszewski (2004:377) observed, “the number and variety of species in an assemblage, *independent of their identities*, provides a means of comparing assemblages from different times and places. This, in turn, can provide information on changes in community structure, as opposed to species membership, over ecological as well as evolutionary timescales” (emphasis added). NTAXA can be the same, say twenty-five, for any two faunas, but those two faunas may not share any taxa, or they may have three or fifteen or twenty-two taxa in common; NTAXA for both will be twenty-five regardless of whether taxa are shared by the two faunas. The same applies to measures of heterogeneity and evenness. Thus one paleoecologist has stated that “in community analysis, communities are described not by their taxonomic content but by their levels of diversity” (Andrews 1996:277); I interpret “levels of diversity” to mean index values of richness, heterogeneity, and/or evenness. A paleozoologist might wish to know how similar two communities (or assemblages) are in terms of shared taxa, and thus a discussion of two commonly used measures of taxonomic similarity is included. First, however, a bit more detail regarding the significance of NTAXA is warranted.

### *Taxonomic Richness*

Flannery (1965, 1969) used the term *broad spectrum* to characterize an early (pre-agricultural) pattern of exploiting numerous kinds of resources and the term *specialization* to label a later, agricultural adaptation in which a small number of resource types were exploited by individual human groups. Cleland (1966, 1976) used the terms “diffuse economy” and “focal economy” to label those that exploited a wide range of resource types and those that used a narrow range of resource types, respectively. Dunnell (1967, 1972) used the terms “extensive” (= diffuse) and “intensive” (= focal) to signify the same resource-exploitation patterns as Cleland. The move was on in ecological anthropology to quantify particular instances of dietary breadth or niche width (Hardesty 1975), and archaeologists kept pace, coining a plethora of terms along the way. Ecologists used the terms “generalized” and “specialized” (e.g., Ricklefs 1979), as did some archaeologists (e.g., Quimby 1960), for what is most fundamentally whether NTAXA is a large value or a small value, respectively. Determination of which of those taxa, plant or animal, comprising the identified assemblage were actually exploited and used by humans is a separate, taphonomic question (Lyman 1994c) and is beyond the scope of discussion.

Modern interest of biologists and ecologists in NTAXA reflects growing concerns over *biodiversity*, a term with a commonsensical meaning but which tends today to imply anthropogenically induced disturbances to biota, especially those that result in a loss of taxonomic variety through extinction (e.g., Pimm and Lawton 1998; Vitousek et al. 1997). NTAXA is one widely understood aspect of biodiversity because it is a fundamental measure and does not require the calculation of a derived value to express it numerically (Gaston 1996). This does not mean that NTAXA values are not difficult to interpret; indeed they are difficult to interpret (Gaston 1996). It is nevertheless not unusual to read about how the richness of one fauna compares to the richness of another. Because the numerical values of NTAXA are whole numbers and can vary from one into the hundreds for any given faunal collection, it is also not unusual to read that one fauna has ten more taxa than another, or one fauna has twice as many taxa as another, or the like. Such statements imply that NTAXA is a ratio scale measurement. In fact it is likely not a ratio scale measurement in biology and ecology for many reasons (Gaston 1996; MacKenzie 2005), some of which are similar to the reasons it is not likely to be a ratio scale measurement in paleozoology.

The value of NTAXA in a particular faunal collection is the easiest mechanically of the four variables of structure and composition to determine. Simply tally how many taxa at some predetermined taxonomic level (family, genus, species) are represented

in a collection; it is a fundamental measure. As noted earlier, the Meier site mammal collection has a taxonomic richness of twenty-six at the genus level, and the complete Cathlapotle mammal collection has a taxonomic richness of twenty-five at the genus level. The Meier mammalian fauna is taxonomically richer than the Cathlapotle mammalian fauna; why the Meier collection is taxonomically richer than the Cathlapotle collection is another matter. It is easy to eliminate one possible answer to this particular *why* question. The Meier collection is smaller ( $\sum \text{NISP} = 5,939$ ) than the Cathlapotle collection ( $\sum \text{NISP} = 6,206$ ), so the difference in richness is likely not a result of sample size; were the Meier collection larger than the Cathlapotle collection, then the difference in richness might be a function of the fact that the Meier collection was larger.

NTAXA can vary over space or through time for any number of reasons. The geographic distributions of species shift daily and seasonally as well as in concert with long-term (multiple year) climatic shifts. Local colonization and extirpation occur, and sometimes an individual waif or vagrant wanders into an area simply by chance. If NTAXA is measured, should it include only resident taxa (ones whose members breed, reproduce, and remain in the area year round) and ignore seasonal immigrants (ones that might breed and reproduce in the area but that are present only part of the year) and vagrant taxa (an individual of which occasionally wanders into the area under study and which may stay there until death but does not reproduce there) (Gaston 1996)? Explicit wording of research questions is the only means to address this question. But there is also a now well-known problem that is a bit more difficult to contend with.

Often, taxonomically richer faunas are larger than less taxonomically rich faunas (e.g., Grayson 1984; Leonard 1989; Sharp 1990). Consider the set of eighteen assemblages from eastern Washington used in earlier analyses. Pertinent data are given in Table 5.1 and graphed in Figure 5.3. As the latter suggests, the two variables,  $\sum \text{NISP}$  and NTAXA (of mammalian genera) are closely correlated ( $r = 0.80$ ,  $p < 0.0001$ ). Knowing the total NISP of any of these collections allows close prediction of the NTAXA in a collection. This is a relationship that has been known for a long time. It is the species-area relationship, and as indicated in Chapter 4, one way to contend with the fact that samples of large size often are taxonomically richer than samples of small size is rarefaction. There are ways other than rarefaction to contend analytically with sample size effects. Use of rarefaction assumes that a sample size influence exists, but fortunately, we need not simply assume such. We can instead determine empirically if a sample size influence exists in any given instance; this involves performing a statistically assisted search for a significant relationship between sample size and richness, such as is exemplified in Figure 5.3. If we find that such a relationship exists,

Table 5.1. Sample size ( $\sum NISP$ ), taxonomic richness ( $S$ ), taxonomic heterogeneity ( $H$ ), taxonomic evenness ( $e$ ), and taxonomic dominance ( $1/D$ ) of mammalian genera in eighteen assemblages from eastern Washington State

Site	$\sum NISP$	$S$	$H$	$e$	$1/D$
45OK18	31	6	1.449	0.809	3.690
45DO204	48	9	1.965	0.894	6.897
45DO273	84	8	1.234	0.594	2.237
45DO243	157	8	1.241	0.597	2.532
45OK2A	366	10	1.342	0.583	2.933
45DO189	415	15	1.362	0.503	2.427
45DO282	426	11	0.894	0.373	1.647
45DO211	474	15	1.490	0.550	2.688
45DO285	491	15	1.812	0.669	3.937
45DO214	536	17	2.059	0.727	5.556
45DO326	640	16	1.985	0.716	4.608
45DO242	673	13	1.260	0.491	2.564
45OK287	807	10	1.310	0.569	2.890
45OK250	1,077	12	1.129	0.454	2.020
45OK4	1,108	15	1.042	0.385	1.835
45OK2	2,574	18	0.861	0.298	1.590
45OK11	3,549	24	1.728	0.544	3.636
45OK258	4,433	21	0.876	0.288	1.608

we need not stop, throw up our hands in dismay, and curse the day. We have several analytical options.

Presuming that there are multiple assemblages, one might compare richness (NTAXA) and sample size ( $\sum NISP$ ) per subset of assemblages to determine if some assemblages display one relationship between the two variables and other assemblages show another relationship (e.g., Grayson 1998; Grayson and Delpech 1998). Figure 5.4 shows two kinds of relationship between the two variables among 13 assemblages at one site, Homestead Cave in western Utah State (Grayson 1998, 2000) (Table 5.2). The oldest three strata (I, II, III) date to a time when climate was moist in what is today a relatively dry area, and all other strata date to times (after 8300  $^{14}C$  yr BP) when it was as dry or drier than today. Ecological theory suggests and empirical data indicate that as moisture increases (either in abundance or effectiveness), primary productivity increases and so too does mammalian taxonomic richness; as moisture decreases, so too does mammalian taxonomic richness (references in Grayson 1998, 2000). More remains of more taxa were accumulated when it was moist and fewer remains of fewer

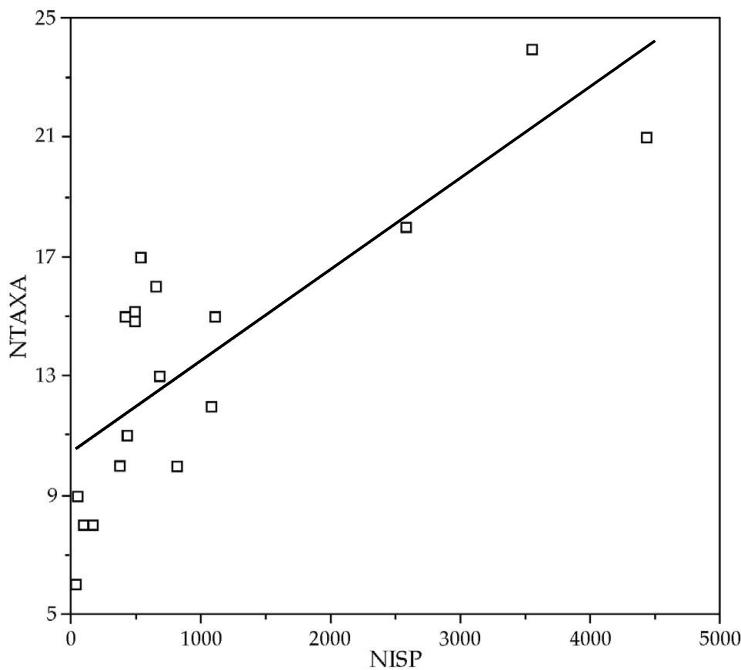


FIGURE 5.3. Relationship between genera richness (NTAXA) and sample size (NISP) in eighteen mammalian faunas from eastern Washington State. Data from Table 5.1.

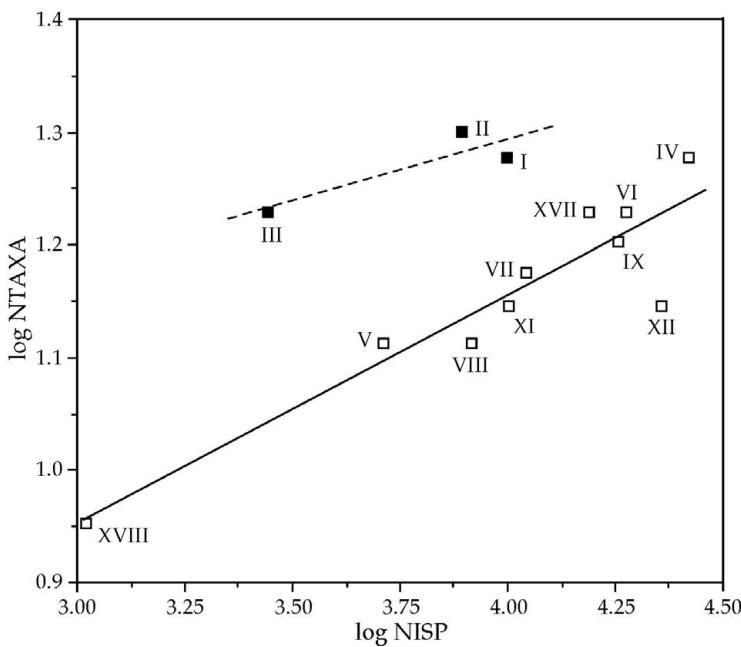


FIGURE 5.4. Relationships between NISP and NTAXA of small mammals per stratum (Roman numerals) at Homestead Cave, Utah (after Grayson 1998). Dashed, best-fit regression line,  $r = 0.88, p = 0.3$ ; solid, best-fit regression line,  $r = 0.92, p = 0.0001$ . Faunal material from strata X and XIII–XVI has not been studied. Data from Table 5.2.

Table 5.2.  $\sum NISP$  and NTAXA for small mammals at Homestead Cave, Utah. Data from Grayson (1998). Faunal remains from strata X and XIII–XVI were not studied

Stratum	$\sum NISP$	NTAXA
XVIII	1047	9
XVII	15,421	17
XII	22,661	14
XI	9,996	14
IX	18,043	16
VIII	8,215	13
VII	11,038	15
VI	18,661	17
V	5,093	13
IV	26,200	19
III	2,774	17
II	7,756	20
I	9,906	19

taxa were accumulated and deposited when it was arid. Despite apparent sample size effects, the mammal assemblages from Homestead Cave appear just as they should in terms of the relationship between NTAXA and climate.

Another example of studying the covariation of sample size and taxonomic richness comes from the Upper Paleolithic rockshelter of Le Flageolet I, France (Grayson and Delpech 1998). The ungulate remains at this site were largely introduced by humans, but interestingly, there is yet another pair of relationships between NTAXA (of ungulates) and  $\sum NISP$  (of ungulates) (Table 5.3). There is no patterned relationship between the associated archaeological culture and which line a particular assemblage of ungulate remains helps define (Figure 5.5). The analysts found no clear indication that the degree of fragmentation was creating the two relationships, and no indication that the differential transport of skeletal parts by bone accumulators had created the two relationships (Grayson and Delpech 1998). They concluded that the difference involved variation in diet breadth, or the width of the niche exploited by the humans that created the assemblages.

There are other ways to compare taxonomic richness values of assemblages of different sizes. Recall from Chapter 4, for example, that the original recognition of the sample-size effect (the species-area relationship) was based on the amount

Table 5.3.  $\sum NISP$  and  $NTAXA$  for ungulates at Le Flageolet I, France. Data from Grayson and Delpech (1998). Strata with  $NISP < 30$  are not included

Stratum	$\sum NISP$	NTAXA
XI	651	6
IX	681	11
VIII	461	9
VII	1,768	10
VI	376	8
V	1,244	7
IV	145	5

of geographic area sampled. Thus, one might compare taxonomic richness with the amount excavated, either the area or volume excavated. Wolff (1975) showed long ago that the greater the volume of sediment searched for faunal remains, the greater the number of taxa found (see Chapter 4). Taxonomic richness increases as the amount of sediment examined increases because as the amount of sediment examined increases, NISP increases (more specimens are recovered, so more taxa

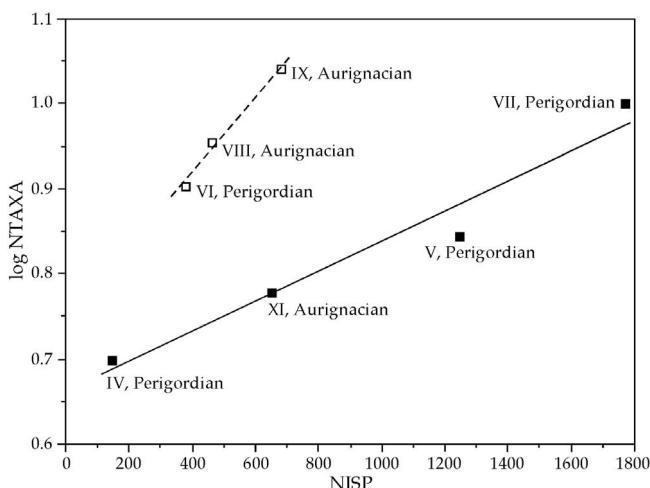


FIGURE 5.5. Relationship between NISP and NTAXA per stratum (Roman numerals) at Le Flageolet I, France (after Grayson and Delpech 1998). Dashed, best-fit regression line,  $r = 0.99$ ,  $p = 0.06$ ; solid, best-fit regression line,  $r = 0.98$ ,  $p = 0.02$ . Some strata omitted as sample sizes are too small for inclusion. Cultural associations for each stratum are indicated. Data from Table 5.3.

are identified). As the amount of sediment examined increases, the amount of area examined increases, which brings us back to the original species-area relationship discovered by botanists.

Regardless of the technique used to gain insight to the structure and composition of a fauna, taxonomic richness is often strongly correlated with sample size. Therefore, the analyst must be ever on the alert for differences in sample size measured as  $\sum$ NISP as a variable that potentially contributes to differences in NTAXA. The analyst should also realize that the possible influence of sample size on all measures of taxonomic diversity (structure and composition) might be disputed, such as when all of a site deposit (all of a single stratum, or all of a site within horizontal and vertical boundaries) has been excavated. In such cases one might argue that a 100-percent sample has been collected and that taxonomic richness cannot be considered to be a function of sample size. This is in some senses true, but it also overlooks two fundamental issues. First, a very small number of sites (or strata within sites other than trivial cases such as the fill of a single intrusive pit) have been totally excavated (meaning that a 100-percent sample has been generated). Second, even if a site is completely excavated, it is likely to be but a portion of some larger cultural system than is found in that single site or only a portion of the taphocoenose, thanatocoenose, or biocoenose. This brings us back to Kintigh's (1984) fundamental dilemma of defining the population one wishes to model with available samples. Only when that population is defined beforehand will we know if we have a 100-percent sample or not, and even then preservation variation and recovery procedures may result in less than complete retrieval (Chapter 4).

### *Taxonomic Composition*

Two faunas can have the same NTAXA, but share anywhere from none to all of the taxa represented. How do we compare faunas in terms of the taxa they hold in common and those taxa that are unique to one or the other? How do we determine if two faunas are similar in taxonomic composition, and how do we determine if fauna A and fauna B are more similar to one another than either is to fauna C? Indices have been designed to answer these questions and to measure just these features (see reviews in Cheetham and Hazel 1969; Henderson and Heron 1977; Janson and Vegelius 1981; Raup and Crick 1979). For unclear reasons these indices have seldom been used by zooarchaeologists (Styles [1981] is a noteworthy exception). It could be a result of benign neglect, or it could be that the influences of varying sample size are a concern. Before considering the latter issue, however, let's consider some

exemplary indices. These are sometimes referred to as binary coefficients, because they summarize and compare presence–absence (nominal scale) data.

One index is the Jaccard index ( $J$ ) [originally, “coefficient of floral community”]. It is calculated as

$$J = 100C/(A + B - C),$$

where  $A$  is the total number of taxa in fauna A,  $B$  is the total number of taxa in fauna B, and  $C$  is the number of taxa common to both A and B. Another index is the Sorenson index ( $S$ ), calculated as

$$S = 100(2C)/(A + B),$$

where the variables are as defined for the Jaccard index. Given how they are calculated, the Jaccard index emphasizes differences in two faunas, and the Sorenson index emphasizes similarities. For example, if  $A = 6$ ,  $B = 6$ , and  $C = 4$ , then  $J = 50$  whereas  $S = 66.7$ . Comparing the Meier site mammalian fauna (A) with the complete Cathlapotle mammalian fauna (B),  $A = 26$ ,  $B = 25$ , and  $C = 20$ . Thus, for these two faunas,  $J = 64.5$  and  $S = 78.4$ . Given that the two faunas fall within the same time period, are  $< 10$  km apart, and occur in virtually identical habitats, it may seem that the indices of faunal similarity should be considerably higher. This is so because, at least with respect to statistical precision sampling (as opposed to discovery sampling; see Chapter 4), at least the Meier site sample seems to be representative because significant increases in its size over the last several years of excavation failed to produce any previously unidentified taxa (Lyman and Ames 2004).

Why the Meier and Cathlapotle assemblages are not more similar and do not share more mammalian genera is an ultimate question. It may have to do with variation in which taxa were accumulated despite similarity of the agents of accumulation; at both sites humans were the most significant accumulation agent. Or, it may actually have to do with a fundamental problem of all such binary coefficients (Raup and Crick 1979). That problem can be illustrated with a pair of Venn diagrams (Figure 5.6). Each of these has been drawn with the Meier and Cathlapotle collections in mind. A total of thirty-one genera are represented by the two collections. One Venn diagram suggests each collection is a sample of those thirty-one genera. The other Venn diagram indicates that each collection is a sample of the total forty mammalian genera (excluding eight genera of bats) that occur in the area today (Johnson and Cassidy 1997). Given that neither zooarchaeological collection has significantly more than two-thirds of those genera, it is perhaps not surprising that the two do not share more taxa. Each site collection represents but a sample of the local biotic community.

Another way to make the point of the preceding paragraph is this: Based on earlier discussions, it should be obvious that sample size ( $= \sum \text{NISP}$ ) will influence binary

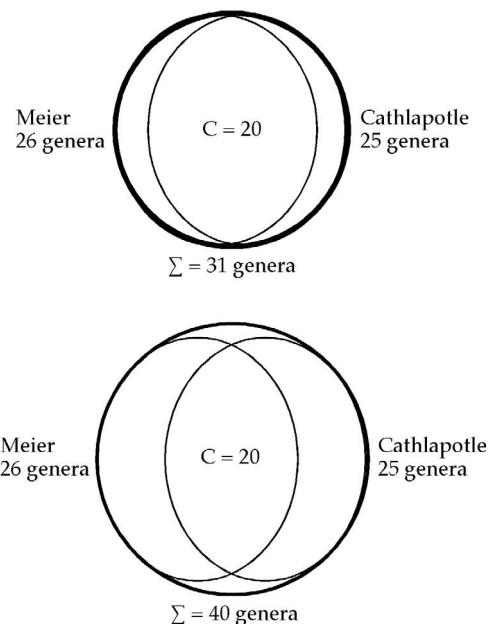


FIGURE 5.6. Two Venn diagrams based on the Meier site and Cathlapotle site collections. Upper diagram suggests each collection is a sample of the thirty-one genera represented by the combined collections. Lower diagram indicates that each collection is a sample of the total forty mammalian genera that occur in the area.

coefficients such as the Jaccard and Sorenson indices. This has been known for decades; the more individuals, the taxonomically richer the sample, so when Paul Jaccard proposed his index he suggested areas of similar size be sampled, but he should have suggested similar numbers of individuals be inspected (Williams 1949). Consider the fact that both the Meier and Cathlapotle collections are samples, and thus even if remains of all forty mammalian genera known in the area today had been accumulated and deposited in site deposits, it is likely that remains of rarely represented taxa would not be recovered. If more of each of those sites had been excavated, and several thousand more NISP had been recovered from each site, it is probable that several of those as yet unidentified genera would occur in those collections. This would not only represent a shift in the sampling design toward a discovery model, but it would also increase the magnitude of both the Jaccard index and the Sorenson index.

Neither the Sorenson index nor the Jaccard index takes advantage of the abundance of taxa. A simple way to assess the similarity of taxonomic abundances of two faunas is to calculate a  $\chi^2$  statistic (e.g., Broughton et al. 2006; Grayson 1991b; Grayson and Delpach 1994). To illustrate this, the NISP data for the collection of faunal remains from eighty-four owl pellets (Table 2.9) is summarized as two chronologically distinct

Table 5.4. NISP per taxon in two chronologically distinct samples of eighty-four owl pellets

Taxon	1999 sample	2000–2001 sample
<i>Sylvilagus</i>	5	0
<i>Reithrodontomys</i>	0	19
<i>Sorex</i>	40	6
<i>Thomomys</i>	52	16
<i>Microtus</i>	302	403
<i>Peromyscus</i>	1,147	119

samples in Table 5.4. Chronological distinction concerns when the pellets were collected.  $\chi^2$  analysis indicates the two samples differ significantly in terms of taxonomic abundances ( $\chi^2 = 586.68, p < 0.0001$ ). The two sets of taxonomic abundances are not correlated (Spearman's  $\rho = 0.6, p > 0.2$ ), which also suggests they may have derived from different populations, but do the abundances of all of the taxa differ significantly between the two samples, or the abundances of just a few of the taxa? To answer this question, adjusted residuals for each cell were calculated (there are six taxa, and 2 years for each, so twelve cells) to determine if any of the observed values were greater, or less than would be expected were the two temporally distinct samples derived from different populations. Basically, the adjusted residual provides a way to determine if the observed and expected values per cell are statistically significantly different or not (see Everitt 1977 for discussion of the statistical method). Expected values (compare with Table 5.4) and interpretations for each cell are given in Table 5.5. Abundances of four taxa are causing the statistically significant difference between the two samples; specimens of *Reithrodontomys*, *Sorex*, *Microtus*, and *Peromyscus* are not randomly distributed between the two chronologically distinct samples. Only *Sylvilagus* and

Table 5.5. Expected values (E) and interpretation (I) of taxonomic abundances in two temporally distinct assemblages of owl pellets. See Table 5.4 for observed values

Taxon	1999 E	2000–2001 E	1999 I	2000–2001 I
<i>Sylvilagus</i>	3.7	1.3	$p > 0.05$	$p > 0.05$
<i>Reithrodontomys</i>	13.9	5.1	$p < 0.05$ , too few	$p < 0.05$ , too many
<i>Sorex</i>	33.7	12.3	$p < 0.05$ , too many	$p < 0.05$ , too few
<i>Thomomys</i>	49.8	18.2	$p > 0.05$	$p > 0.05$
<i>Microtus</i>	516.8	188.2	$p < 0.05$ , too few	$p < 0.05$ , too many
<i>Peromyscus</i>	928.0	338.0	$p < 0.05$ , too many	$p < 0.05$ , too few

*Thomomys* occur in the two samples in abundances that are not unexpected; abundances of these two taxa suggest the temporally distinct samples were drawn from the same population.

Some research has suggested that the Sorenson index provides a better estimate of similarity than Jaccard's index (Magurran 1988:96). Not surprisingly, ecologists designed a version of Sorenson's index to take account of variation in taxonomic abundances. That index, Sorenson's quantitative index, is calculated as

$$S_q = 2cN/(AN + BN),$$

where  $AN$  is the total frequency of organisms (all taxa summed) in fauna A,  $BN$  is the total frequency of organisms in fauna B, and  $cN$  is the sum of the lesser of the two abundances of taxa shared by the two assemblages. Using the data in Tables 4.2 and 4.3 for Meier and Cathlapotle mammalian genera,  $AN$  (Meier) = 6421,  $BN$  (Cathlapotle) = 6,937, and  $cN$  = 4,358 ( $3\text{Scapanus} + 7\text{Aplodontia} + 342\text{Castor} + 5\text{Peromyscus} + 68\text{Microtus} + 106\text{Ondatra} + 39\text{Canis} + 5\text{Vulpes} + 102\text{Ursus} + 207\text{Procyon} + 2\text{Martes} + 29\text{Mustela} + 3\text{Mephitis} + 51\text{Lutra} + 9\text{Felis} + 26\text{Lynx} + 43\text{Phoca} + 935\text{Cervus} + 2376\text{Odocoileus}$ ). Thus,  $S_q = 2(4358)/(6421 + 6973) = 8716/13,394 = 0.651$ , or 65.1. Recall that the (nonquantitative) Sorenson's index was 78.4. Thus, regardless of which index of similarity is used, the faunas seem fairly similar, though less so when the abundances of taxa are included than when they are ignored.

A simple way to show similarities and differences between two faunas in terms of shared taxa, unique taxa, and taxonomic abundances, is to generate a bivariate scatterplot. Figure 5.7 shows relative (percentage) abundances of those taxa from Meier and Cathlapotle represented by NISP < 200 at both sites. Notice that were the relative abundances of taxa equivalent at the two sites, the points would fall close to the diagonal line; the more equal the relative abundances, the closer to the diagonal the points would fall. Note as well that more of the points fall on the Meier side of the diagonal. This suggests that those taxa are relatively more abundant at Meier than they are at Cathlapotle. Such a graph takes advantage of abundance data in a visual way. Ecologists are working to develop versions of the Jaccard and Sorenson indices that also take advantage of abundance data (e.g., Chao et al. 2005), but these are beyond the scope of the discussion here.

That the binary coefficients designed to measure taxonomic similarities of faunal collections have been largely ignored by zooarchaeologists is likely a good thing. Those coefficients are heavily influenced by the sample sizes ( $= \sum\text{NISP}$ ) of the compared collections because taxonomic richness is significantly influenced by sample size (Chao et al. 2005). Again, one might use rarefaction in an effort to control sample-size effects, and that is what some paleozoologists have done (e.g., Barnosky et al. 2005;

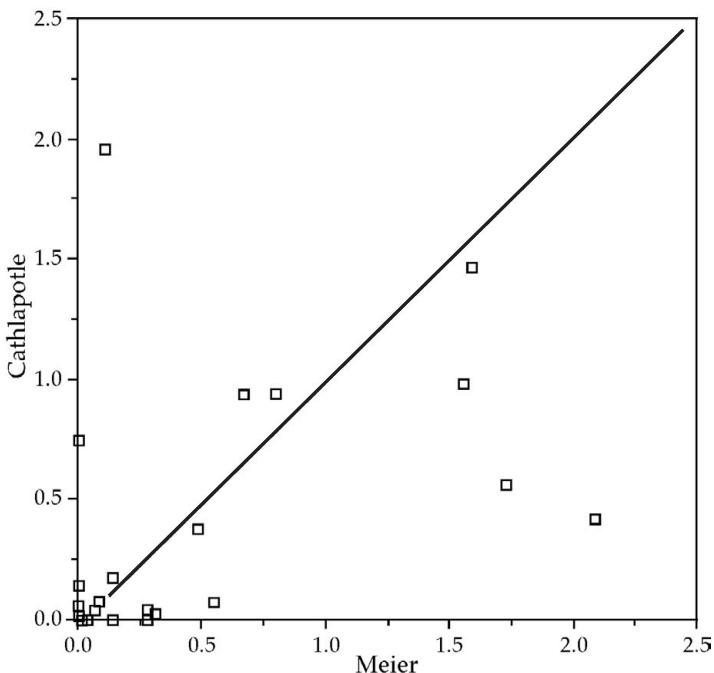


FIGURE 5.7. Bivariate scatterplot of relative (percentage) abundances of mammalian genera at the Meier site and Cathlapotle. Only genera for which NISP < 200 are plotted. Diagonal line is shown for reference.

Byrd 1997). This could be a good thing, but it is perhaps unwise for the simple reason that as was noted more than 20 years ago, the rarefaction procedure was designed to be used with quantitative units that are statistically independent of one another (Grayson 1984:152). Those units are also ratio scale values. NISP tallies comprise units that are probably statistically interdependent and that are also typically at best ordinal scale values. Given these facts, should one choose to perform a rarefaction analysis using NISP values, the results should be interpreted in at most ordinal scale terms. An example will make this clear.

A rarefaction curve based on the eighteen assemblages listed in Table 5.1 constructed using Holland's (2005) *Analytical Rarefaction* is shown in Figure 5.8 and suggests that, given a total NTAXA of twenty-eight for the area represented by those eighteen collections, none of the collections contains all twenty-eight taxa, most collections contain very few taxa, but none contain too few for their size, and four (45DO214, 45DO326, 45DO211, 45DO285) of the eighteen collections seem to contain more taxa than they should given their size ( $\sum$ NISP). The rarefaction curve

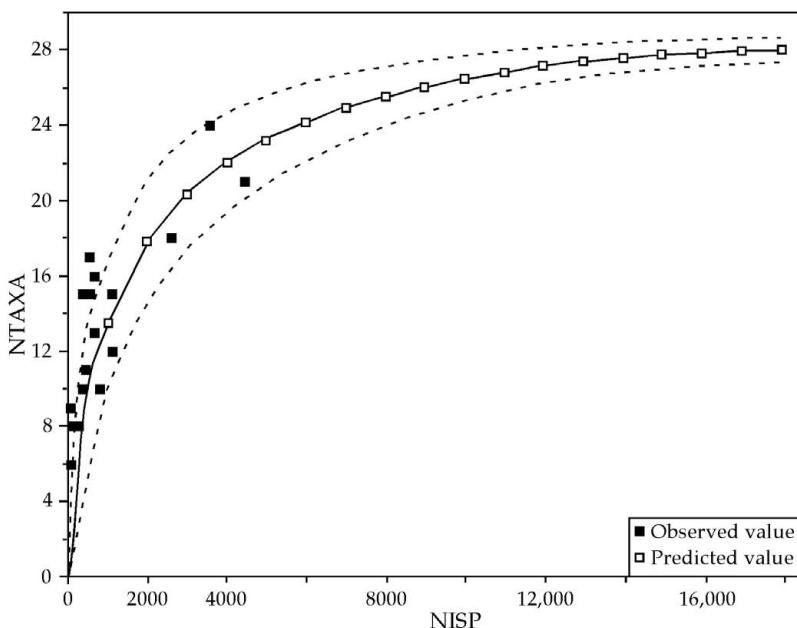


FIGURE 5.8. Rarefaction analysis of eighteen assemblages of mammal remains from eastern Washington State using Holland's (2005) *Analytical Rarefaction*. Data from Table 5.1.

thus reveals something we didn't know before because it presents the data in a unique, interpolated way. If I were analyzing these collections, I would try to determine why four of the collections were unexpectedly taxonomically rich; perhaps they are temporally unique, functionally/behaviorally unique, or located in a particular microenvironment.

Were one to perform a rarefaction analysis like that shown in Figure 5.8, one should first determine if those assemblages are nested. Recall from Chapter 4 that in a series of perfectly nested faunas, successively smaller faunas will have fewer of the taxa represented in those faunas than are successively larger, and larger faunas will have all those taxa represented in smaller faunas plus additional taxa. The interpretive assumption is that nested faunas all derive from the same parent population. There are ways to test the degree of nestedness of faunas. That has been done with the faunas in Figure 5.8; consider Figure 4.12, which shows that the nestedness "temperature" for this set of eighteen faunas is  $18.23^\circ$ , meaning the faunas are relatively strongly nested. The rarefaction analysis in Figure 5.8 thus seems reasonable, if one is willing to allow an unknown degree of skeletal specimen interdependence and, thus, allow a bit of statistical sloppiness.

### *Taxonomic Heterogeneity*

Several indices have been developed to measure taxonomic heterogeneity. Paleozoologists have tended to use only two of these, although there are several different ones that are occasionally mentioned (e.g., Andrews 1996). By far the most popular one among zooarchaeologists is the Shannon–Wiener index, sometimes referred to as the Shannon index. It generally varies between 1.5 and 3.5 (Magurran 1988:35); larger values signify greater heterogeneity. The Shannon index is calculated as:

$$H = - \sum P_i (\ln P_i),$$

where  $P_i$  is the proportion (P) of taxon  $i$  in the assemblage. The proportion (sometimes referred to as “importance”) of each taxon in the collection is multiplied by the natural log of that proportion. Because proportions are  $< 1$ , transforming those values to natural logs results in a negative sign. Values of the products of the multiplications are summed, and then converted from a negative value to a positive value by the negative or “–” sign in front of the summation ( $\sum$ ) sign.

Let’s say we want to determine the taxonomic heterogeneity (at the genus level) of the total Meier site mammal collection (Table 4.2). The data and mathematical steps for calculating the value of the Shannon–Wiener index are summarized in Table 5.6. NTAXA for this collection is 26. The heterogeneity index is 1.556, suggesting the total Meier site mammal collection is somewhat heterogeneous. For comparative purposes, consider the fact that the Shannon–Wiener heterogeneity index for the total Cathlapotle collection, without distinction of the Precontact and Postcontact assemblages (Table 4.3), has a value of 1.487, indicating that the heterogeneity of the Cathlapotle collection is a bit less than that of the Meier collection. One contributing factor here is that the Meier collection, with twenty-six taxa, is taxonomically richer than the Cathlapotle collection, which contains remains of only twenty-four taxa. Does a difference in the evenness of the two assemblages also contribute to the difference in heterogeneity? To answer that question requires calculation of an evenness index for each collection.

Because heterogeneity is a function of taxonomic richness and evenness, it is possible that heterogeneity will also be a function of sample size (e.g., Grayson 1981b). Thus, if one wishes to measure heterogeneity and compare that variable across several different samples, it is advisable to determine if there is any relationship between the measures of heterogeneity and  $\sum$ NISP for a set of samples. Once again, consider the eighteen assemblages from eastern Washington State (Table 5.1). The relationship between sample size per site and heterogeneity per site is, in the case of these eighteen

Table 5.6. Derivation of the Shannon–Wiener index of heterogeneity for the Meier site (original data from Table 4.2). Logs are natural logarithms

Taxon	NISP	Proportion ( $p$ )	Log of $p$	$p(\log p)$	Running sum
<i>Scapanus</i>	18	0.00280	−5.878	−0.016	−0.016
<i>Sylvilagus</i>	18	0.00280	−5.878	−0.016	−0.032
<i>Aplodontia</i>	7	0.00109	−6.822	−0.007436	−0.039
<i>Tamias</i>	1	0.00016	−8.740	−0.001398	−0.041
<i>Tamiasciurus</i>	2	0.00031	−8.079	−0.002504	−0.043
<i>Thomomys</i>	9	0.00140	−6.571	−0.009199	−0.053
<i>Castor</i>	342	0.05326	−2.933	−0.156	−0.209
<i>Peromyscus</i>	35	0.00545	−5.212	−0.028	−0.237
<i>Rattus</i>	1	0.00016	−8.740	−0.001398	−0.238
<i>Neotoma</i>	1	0.00016	−8.740	−0.001398	−0.239
<i>Microtus</i>	100	0.01557	−4.162	−0.065	−0.304
<i>Ondatra</i>	374	0.05825	−2.843	−0.166	−0.470
<i>Erethizon</i>	1	0.00016	−8.740	−0.001398	−0.472
<i>Canis</i>	111	0.01729	−4.058	−0.070	−0.542
<i>Vulpes</i>	5	0.00078	−7.156	−0.00582	−0.547
<i>Ursus</i>	102	0.01589	−4.142	−0.066	−0.613
<i>Procyon</i>	287	0.04470	−3.108	−0.139	−0.752
<i>Martes</i>	20	0.00311	−5.773	−0.018	−0.770
<i>Mustela</i>	134	0.02087	−3.869	−0.081	−0.851
<i>Mephitis</i>	4	0.00062	−7.386	−0.004579	−0.856
<i>Lutra</i>	51	0.00794	−4.836	−0.038	−0.894
<i>Puma</i>	9	0.00140	−6.571	−0.009199	−0.903
<i>Lynx</i>	31	0.00483	−5.333	−0.026	−0.929
<i>Phoca</i>	43	0.00670	−5.006	−0.034	−0.963
<i>Cervus</i>	935	0.14562	−1.927	−0.281	−1.244
<i>Odocoileus</i>	3,780	0.58869	−0.530	−0.312	−1.556
$\sum =$	6,421	1.0	—	—	$-\sum = 1.556$

assemblages, not statistically significant, suggesting that variation in heterogeneity is not being driven by variation in sample size (Figure 5.9). Given the insignificant correlation between sample size and heterogeneity, we might feel safe in comparing heterogeneities across these eighteen assemblages were it not for the fact that sample size and richness are strongly correlated among them (Figure 5.3), and the fact that as richness increases so too does heterogeneity. Only in those instances in which there is no such correlation (between sample size and richness and between sample size and heterogeneity) might one make statements about variation between assemblages

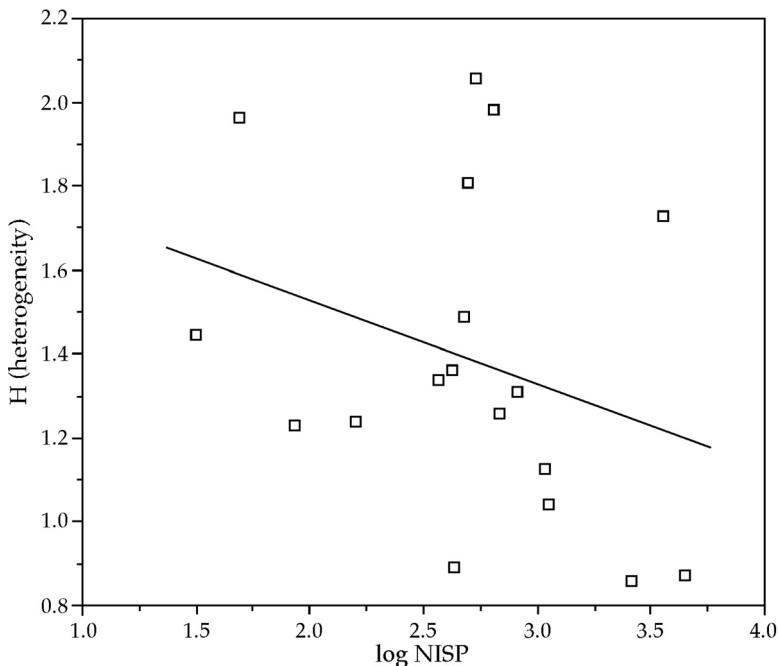


FIGURE 5.9. The relationship between taxonomic heterogeneity ( $H$ ) and sample size (NISP) in eighteen assemblages of mammal remains from eastern Washington State. Best-fit regression line is insignificant ( $r = -0.30$ ,  $p > 0.2$ ). Data from Table 5.1.

and taxonomic heterogeneity. Otherwise, variation in heterogeneity may be a result of variation in sample size across the compared collections.

Notice that I said “might” and “may” with respect to the possible influence of sample size on heterogeneity. Heterogeneity is a summary measure of the *relative* (proportional) abundances of taxa. There is, therefore, a potential problem with the analysis attending Figure 5.9. Specifically, the problem concerns the fact that *relative* abundances are involved. That problem can be circumvented by using a statistical test other than a form of correlation analysis. That test will be introduced after evenness is discussed.

#### Taxonomic Evenness

How are individuals distributed across taxa? Are all taxa represented by about the same number of individuals (an *even* fauna), or do some taxa contain many individuals whereas other taxa contain very few individuals (an *uneven* fauna)? To answer such questions beyond comparing histograms showing the distribution of individuals across taxa in each fauna (the histograms may not be visually distinguishable;

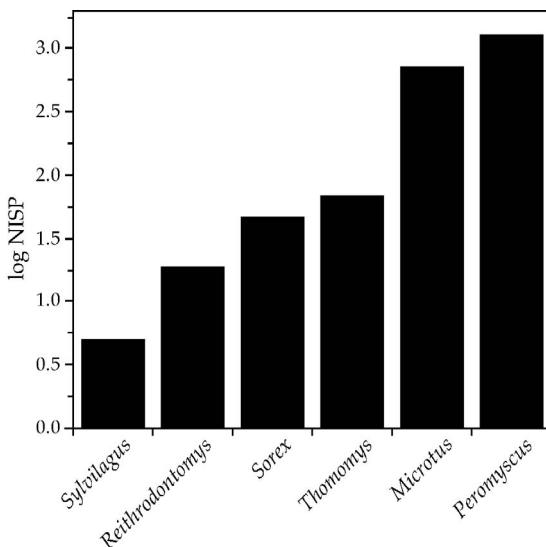


FIGURE 5.10. Frequency distribution of NISP values across six mammalian genera in a collection of owl pellets. Data from Table 2.9.

Figure 5.2), an index of evenness can be calculated. One index that is regularly used to measure evenness quantitatively requires the Shannon–Wiener heterogeneity index, which is divided by the log of NTAXA or richness (Magurran 1988). Thus, evenness is calculated as:

$$e = H / \ln S,$$

where  $H$  is the Shannon–Wiener heterogeneity index and  $S$  is taxonomic richness. The lower the value of  $e$ , known as the Shannon index of evenness (Magurran 1988), the less even the assemblage. The Shannon index,  $e$ , is constrained to fall between 0 and 1, with a value of 1 indicating an even fauna or that all taxa are equally abundant (Magurran 1988).

Consider the assemblage of mammals represented in the previously mentioned owl pellet collection and described in Table 2.8. NTAXA ( $S$ ) is 6, using the NISP values for this collection the Shannon–Wiener heterogeneity index ( $H$ ) is 0.922, and the Shannon index of evenness ( $e$ ) is  $(0.922 / 1.792) = 0.515$ . Inspection of Figure 5.10, which shows the frequency distribution of NISP across the taxa, suggests that indeed, this fauna is not very even, just as its calculated  $e$  value suggests.

Recall the question posed above after the Shannon–Wiener heterogeneity indices for the Meier collection and for the Cathlapotle collection were calculated – the Meier collection was more heterogeneous in part because it was taxonomically richer than the Cathlapotle collection, but did a difference in the evenness of the two faunas also

contribute to that difference in heterogeneity? For Meier,  $e = 1.556/3.258$  (where 3.258 is the natural log of 26), so  $e = 0.4776$ ; for Cathlapotle,  $e = 1.487/3.178$  (where 3.178 is the natural log of 24), so  $e = 0.4679$ . The evenness index  $e$  varies between 0 and 1; if  $e = 1$ , then all taxa are equally abundant (specimens are equally distributed across taxa). So, the Meier collection seems to be slightly more even than the Cathlapotle collection—identified specimens are a bit more equally distributed across the twenty-six taxa of the Meier collection than identified specimens are distributed across the twenty-four taxa of the Cathlapotle collection. Thus, not only does the greater richness of the Meier collection contribute to greater heterogeneity, so too does the greater evenness of the Meier collection contribute to its heterogeneity being greater than that evident in the Cathlapotle collection.

As with richness and heterogeneity, there are cases when taxonomic evenness seems to be driven by sample size, though there are also cases where there is no such relationship (e.g., Grayson and Delpech 1998; Grayson et al. 2001; Nagaoka 2001, 2002). Until a few years ago, to determine whether or not evenness was driven by sample size, one measured the strength of the relationship between the two variables. Consider the eighteen assemblages of mammal remains from eastern Washington State (Table 5.1). Sample size and evenness values among these eighteen assemblages are strongly correlated (Figure 5.11); about 53 percent of the variation in evenness is explained by variation in NISP ( $r = 0.73, p < 0.002$ ). On the basis of that correlation, the paleozoologist would have previously concluded that it would be ill-advised to suggest variation in evenness across these assemblages was due to some ecological, environmental, or human behavioral variable rather than simple difference in sample size. There is now a better way to search for sample size effects on measures of relative abundance that will be described shortly. First, another measure of evenness needs to be introduced.

Another index of evenness occasionally used by zooarchaeologists is the reciprocal of Simpson's index (Grayson and Delpech 2002; James 1990; Jones 2004; Schmitt and Lupo 1995; Wolverton 2005). If one assumes that the population is infinitely large, the index is calculated as

$$1 / \sum p_i^2,$$

where  $p_i$  is the proportional abundance of taxon  $i$  in the total collection (Magurran 1988:39). However, because the population of organisms in a community, and the sample of remains in a deposit, are finite, it is appropriate to calculate Simpson's index as:

$$\sum n_i [n_i - 1] / N[N - 1],$$

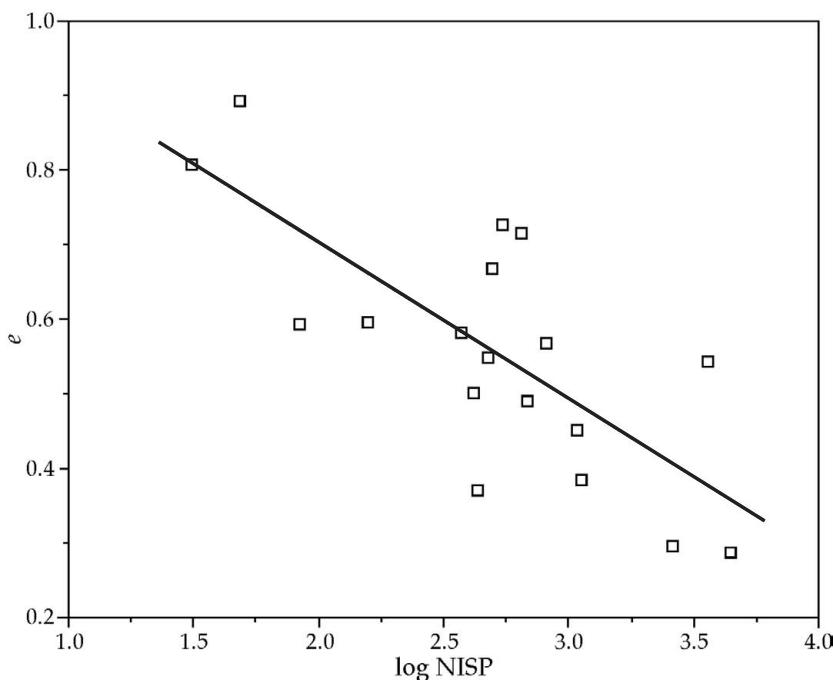


FIGURE 5.11. Relationship between taxonomic evenness ( $e$ ) and sample size (NISP) in eighteen assemblages of mammal remains from eastern Washington State. Best-fit regression line is significant ( $r = -0.73$ ,  $p = 0.0005$ ). Data from Table 5.1.

where  $n_i$  is the number of specimens of the  $i$ th taxon, and  $N$  is the total number of specimens of all taxa (Magurran 1988). The more evenly distributed individuals (or specimens) are across taxa, the larger the value of this index. Simpson's index is known as  $D$  because it is more sensitive to the dominance of an assemblage by a single taxon than is  $e$  and  $D$  is also less sensitive to taxonomic richness (Magurran 1988). The reciprocal or inverse of Simpson's index is used by ecologists and paleozoologists because the lower the value of the reciprocal, the more the assemblage is dominated by a single taxon, or the less evenly individuals are distributed across all taxa in the assemblage. Thus, as  $1/D$  decreases, the more an assemblage is dominated by a single taxon. The reciprocal of Simpson's index for the owl pellet assemblage (Table 2.9) is 1.314, suggesting that the assemblage is fairly strongly dominated by a single taxon (in this case, *Peromyscus*), just as we would conclude by simple inspection of Figure 5.10.

The inverse of Simpson's index of dominance, or  $1/D$ , for each of the eighteen assemblages of mammalian genera from eastern Washington State is given in Table 5.1. Those index values are only weakly correlated with the NISP per sample size ( $r = -0.43$ ,  $0.1 > p > 0.05$ ), as might be surmised from the wide scatter of points around

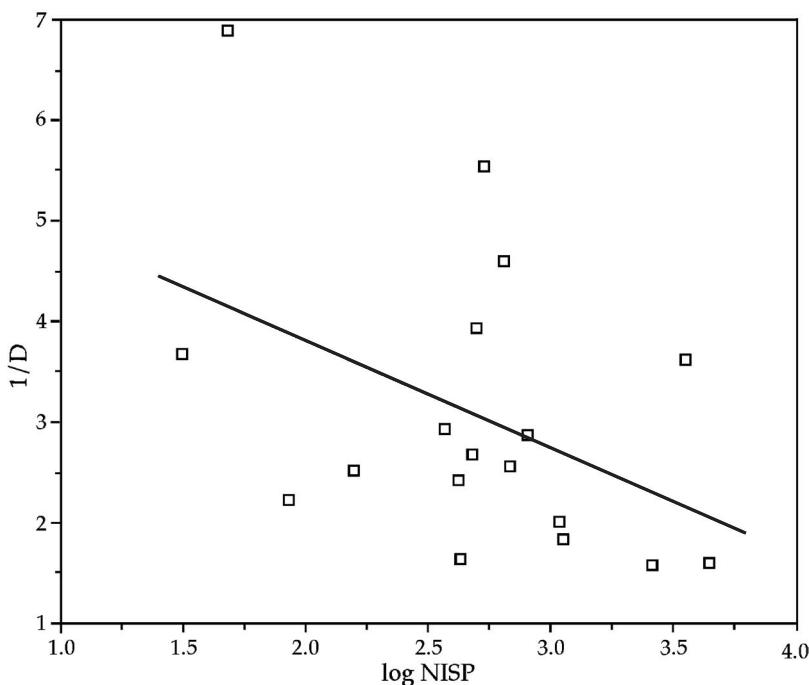


FIGURE 5.12. Relationship between sample size and the reciprocal of Simpson's dominance index ( $1/D$ ) in eighteen assemblages of mammal remains from eastern Washington State. Best-fit regression line is weakly significant ( $r = -0.43, 0.1 > p > 0.05$ ). Data from Table 5.1.

the simple best-fit regression line (Figure 5.12). Thus, one would likely be tempted to conclude that taxonomic evenness or dominance in these eighteen assemblages does not seem to be significantly influenced by sample size and thus one could discuss differences in evenness or dominance in terms of ecological variation in site settings or variation among the taphonomic agents that accumulated the remains or the like. But again, a correlation has been sought between sample size ( $\sum\text{NISP}$ ) per assemblage and a measure of relative abundance (proportions). It is now time to discuss that issue.

### *Discussion*

Simple regression analysis such as is illustrated in Figure 5.9 is a reasonable way to search for sample size effects if absolute measures of sample size are used, such as when one seeks to determine if there is a relationship between NTAXA and  $\sum\text{NISP}$ . What

Table 5.7. Total NISP of mammals, NISP of deer (*Odocoileus spp.*), and relative (%) abundance of deer in eighteen assemblages from eastern Washington State

Site	$\sum$ NISP	Deer NISP	% Deer
45OK18	31	0	0
45DO204	48	5	10.4
45DO273	84	6	7.1
45DO243	157	34	21.7
45OK2A	366	87	23.8
45DO189	415	251	51.8
45DO282	426	4	0.9
45DO211	474	45	9.5
45DO285	491	33	6.7
45DO214	536	184	34.3
45DO326	640	120	18.8
45DO242	673	368	54.7
45OK287	807	252	31.2
45OK250	1,077	738	68.5
45OK4	1,108	803	72.5
45OK2	2,574	2,021	78.5
45OK11	3,549	1,668	47.0
45OK258	4,433	3,458	78.0

about those cases when relative abundances (percentages or proportions) of one or more taxa are of interest? Perhaps the paleozoologist wishes to know if the percentages of a taxon in a series of assemblages increase or decrease in a patterned manner (over time if the assemblages are chronologically distinct, or over space if the assemblages are arrayed over an expanse of geographic space). Consider the relative abundance of deer in the eighteen assemblages from eastern Washington State (Table 5.7). The percentage of each assemblage's  $\sum$ NISP representing deer is strongly correlated with the  $\sum$ NISP per assemblage (Spearman's  $\rho = 0.763$ ,  $p = 0.0002$ ), suggesting there may be a sample-size effect driving the trend. Grayson (1984:126–127) suggested an empirical way to contend with cases when one finds a correlation between the relative abundance of one or more taxa and NISP across multiple samples: “Remove assemblages in order of increasing sample size until the correlation between sample size and relative abundance is no longer significant.” Doing so with the eighteen eastern Washington State assemblages results in elimination of the ten smallest

assemblages – those with  $\sum \text{NISP} < 600$ . When those ten assemblages are eliminated, the correlation between the relative abundance of deer and total assemblage size is no longer significant ( $\rho = 0.667$ ,  $p = 0.08$ ).

There is a way to evaluate whether relative abundances are influenced by sample size without eliminating assemblages. And it is more statistically sensitive to detecting true sample size effects than the correlation technique when percentages or proportions are used as measures of abundance. Cannon (2000, 2001) pointed out that correlating (spatial or temporal) trends in *relative* abundances of taxa with sample size ( $\sum \text{NISP}$ ) is statistically not the best way to search for sample size effects. This is so because relative abundances do not register whether sample sizes are five or five thousand. Statistically, noting the difference between an absolute tally of five and an absolute tally of five thousand is quite different than saying each comprises 5 percent of the total collection (100 and 100,000, respectively). Small samples make ruling out sample size effects difficult; their effect on correlations may simply be due to sampling error rather than any accurate reflection of abundance. Rare phenomena are particularly difficult to inventory – they will be absent from collections – unless the samples are large. If the abundances of several rare taxa are not quite equal in the target population, but samples are small, the true relative abundances of those rare taxa likely will not be accurately reflected by small samples (Grayson 1978a, 1979, 1984). A correlation between the relative abundance of a taxon and  $\sum \text{NISP}$  across multiple assemblages may be driven by small samples simply because of sampling error inherent in those assemblages.

Using simulated samples drawn from fictional but known populations, Cannon (2001) showed that absolutely small samples of populations that have trends in the relative abundance of a taxon often display no trends, and also that absolutely small samples of populations that have no trends in the relative abundance of a taxon often display trends. Thus, using the regression approach to search for sample size effects (as with how the data in Table 5.7 were examined) may lead to commission of a Type I error (rejecting a true null hypothesis that there is no true trend in relative abundances when in fact there is no trend) or commission of a Type II error (accepting a false null hypothesis that there is no true trend in abundances when in fact there is a trend). In both cases, the null hypothesis is that no trend is present, but sampling error has produced samples that are not representative of the population. Furthermore, the presence of a significant correlation coefficient is interpreted as indicating that sample size is the source of the correlation, and the absence of a significant correlation coefficient is interpreted as indicating that sample size is not the source of the correlation. As Cannon (2001:185) astutely observed, the

first interpretation at best rests on an incomplete understanding of the relationship between relative abundances and sample size; the second interpretation presumes sample sizes are sufficiently large to warrant confidence but in fact may be too small.

Cannon (2000, 2001) suggested that rather than regression analysis or calculation of a correlation coefficient, a different statistical test be used to ascertain if sample size effects plague an analysis of relative abundances. Cochran's test of linear trends is a form of  $\chi^2$  analysis that tests for trends among multiple rank-ordered samples (Zar 1996:562–565). As Cannon (2000:332) notes, Cochran's test is constructed such that “significant trends will not be found when samples are so small that random error cannot be ruled out at a specified confidence level as the cause of differences in relative abundance between samples.” Cochran's linear test seeks trends (either across space or over time) in relative abundance in such a way as to more directly take absolute sample size into account than correlation-based analyses. One first calculates a standard  $\chi^2$  statistic, and then determines how much of that statistic is the result of a linear trend; if the latter is sufficiently (statistically significant) large, then one concludes that there is indeed a linear trend in the data *independent of sample size*.

Let us return yet again to the eighteen assemblages from eastern Washington State and the relative abundance of deer remains to determine if there is a linear trend in the relative abundance of deer or not (Table 5.7). The overall  $\chi^2$  statistic is large and significant ( $\chi^2 = 4196.6, p < 0.0001$ ), suggesting there is a significant association between sample size and frequency of deer remains. The  $\chi^2$  statistic for a linear trend is also significant ( $\chi^2 = 2239.3, p < 0.0001$ ), suggesting there is significant trend in the abundance of deer remains across the eighteen assemblages regardless of the sizes of those assemblages. Figure 5.13 suggests that there is indeed a trend in relative abundance of deer across the eighteen assemblages ( $r = 0.79, p < 0.001$ ). The problem is that Figure 5.13 gives no indication of possible sample size influences on those relative abundances. Cochran's test for linear trends does just that, though due to its relatively recent introduction to paleozoology (Cannon 2000, 2001), it has as yet seldom been used (e.g., Cannon 2003; Nagaoka 2005a; Wolverton 2005).

There are several other, nonstatistical points to keep in mind regarding whether or not sample size influences seem to exist. One is that taxonomic richness, heterogeneity, and evenness (regardless of the exact index calculated) were designed for extant ecological communities (e.g., Jones 2004), and they were designed to use tallies of individual (statistically independent) organisms as the quantitative units (Grayson 1984). Because it is likely that NISP values are not statistically independent tallies per taxon, cautious interpretation of paleozoological values for richness, heterogeneity,

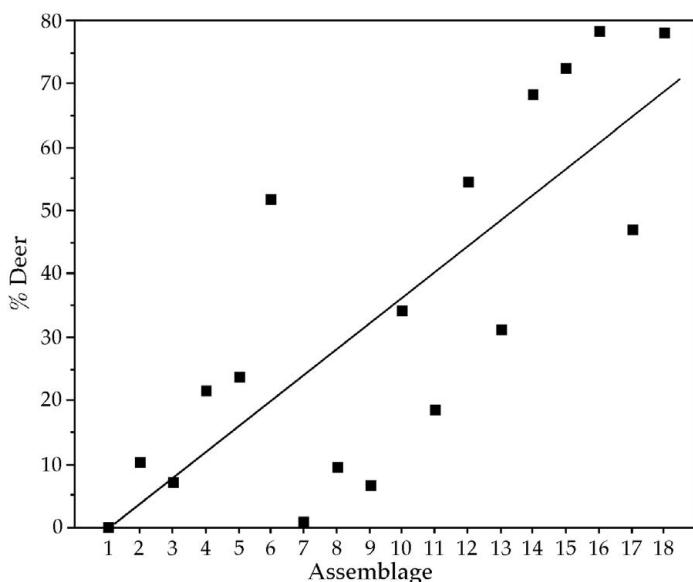


FIGURE 5.13. Percentage abundance of deer in eighteen assemblages from eastern Washington State.

and evenness should be foremost in one's mind. Do not be misled into thinking about these variables as if they are ratio scale variables; chances are good that they are not, and chances are fair that they may not even be ordinal scale variables. If it can be shown that taxonomic abundance data based on NISP are ordinal scale variables (see Chapter 2 for description of a method), then it can be argued that the index values are also ordinal scale values.

Another point made by Nagaoka (2001, 2002) about evenness extends to heterogeneity. She noted that evenness does not take into account the position of taxa in a rank ordered (based on abundance) set of taxa. Using her example, taxon A may comprise 80 percent and taxon B 20 percent of assemblage I, but taxon B comprises 20 percent and taxon A 80 percent of assemblage II. Evenness and heterogeneity indices will not register these obvious differences, so inspection of how much each taxon is contributing may be required if something other than simply an index of faunal structure (is the fauna even or uneven, heterogeneous or homogeneous) is desired. Jones (2004) adds that evenness will be influenced by NTAXA (as will heterogeneity). The important point therefore is that evenness and heterogeneity cannot be considered independently of richness or of each other. One might choose to focus analysis on one variable, but the other variables (with the possible exception of richness) will likely require examination in order to correctly interpret the target variable.

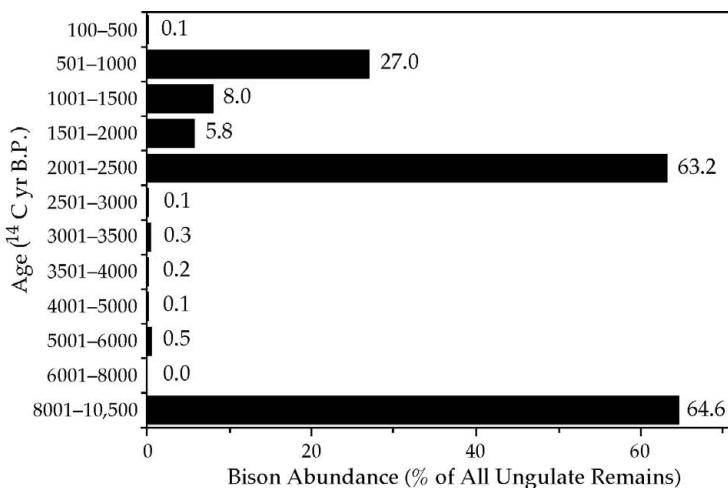


FIGURE 5.14. Abundance of bison remains relative to abundance of all ungulate remains over the past 10,500 ( $C_{14}$ ) years in eastern Washington State.

#### TRENDS IN TAXONOMIC ABUNDANCES

For nearly as long as paleozoologists have measured taxonomic abundances evident in the collections of remains they study, they have also tracked those abundances through time and across space. Sometimes that tracking has involved tables of numbers, each column representing a taxon, each row a stratum if the analyst was interested in temporal variation in abundances. If the analyst was interested in variation in taxonomic abundances across space, each row could be a geographically distinct collection locality whereas each column was a distinct taxon. Tables of numbers presented data, but if more than a few rows and columns were included, such tables are difficult to interpret visually. A reasonable alternative, then, was to construct a graph displaying taxonomic abundances across time or space, although raw data might not be included.

Both tables of numbers and graphs of taxonomic abundances are still constructed by paleozoologists because they are useful analytical techniques. The graph in Figure 5.14 is based on data in Table 5.8 (from Lyman 2004b). The graph shows the relative abundance of bison (*Bison* spp.) remains in eastern Washington State more or less by 500-year bins since the terminal Pleistocene; several bins are lumped to simplify the graph. The graph is significant for several reasons. First, it shows the abundance of bison remains relative to the abundance of remains of all other ungulates in the paleozoological record. There is no correlation between the relative abundance of

Table 5.8. *Frequencies (NISP) of bison and nonbison ungulates per time period in ninety-one assemblages from eastern Washington State*

Time period (yr BP)	Bison ( <i>Bison</i> sp.)	Nonbison ungulates	Total
100–500	4	4,196	4,200
501–1,000	107	3,838	3,945
1,001–1,500	131	1,512	1,643
1,501–2,000	73	1,182	1,255
2,001–2,500	375	218	593
2,501–3,000	4	3,851	3,855
3,001–3,500	5	1,639	1,644
3,501–4,000	1	414	415
4,001–5,000	1	776	777
5,001–6,000	10	2,111	2,121
6,001–8,000	0	405	405
8,001–10,500	186	102	288

bison and the number of assemblages per period ( $p > 0.5$ ); there is no correlation between the total ungulate NISP and bison NISP per temporal period ( $p > 0.5$ ). The abundance of bison remains does not seem to be a function of sampling intensity or sample size.  $\chi^2$  analysis indicates there are significant differences in the relative abundances of bison remains in the samples ( $\chi^2 = 8291.5, p < 0.0001$ ); the test for linear trends indicates there is a trend in the abundance of bison remains ( $\chi^2$  trend = 109.55,  $p < 0.0001$ ). The data indicate an increase in bison over time. Grass was least abundant in the geographic area of concern between 8,000 and 4,000  $^{14}\text{C}$  years ago, consistent with the few remains of bison; bison eat mostly grass. Bison likely immigrated from Montana and Wyoming through southern Idaho, 2,500 years ago or later, when fire frequencies increased (based on the amount of charcoal deposited in lake sediments) and made the immigration route more hospitable (Lyman 2004b). The graph in Figure 5.14 shows the history of the frequency of bison remains in eastern Washington clearly. Similar graphs have been around in paleozoology for several decades (e.g., Ziegler 1973; Stiner et al. 2000). The graph in Figure 5.14 plots only the relative abundance of bison remains; most other graphs plot multiple taxa simultaneously and thus such graphs can sometimes be difficult to interpret.

Another way to graph taxonomic abundances that has seen some recent popularity is of more deductive derivation. Popularized in the early 1990s, a group of paleozoologists, many of whom work in the western United States, calculated indices of

taxonomic abundance and plotted those index values against (usually) time or (less often) space. The seminal work was Bayham's (1979) plotting of the ratio of abundances of large to small animals. His reasoning in doing so was that large animals (grouped by taxon) were more efficiently exploited (cost less in terms of energy expended relative to energy earned) than were small animals given tenets of foraging theory (see Stephens and Krebs [1986] for a theoretical statement). Bayham's notion and his method were refined by Broughton (1994a, 1994b, 1999; see also James 1990) in a series of studies on mammalian, piscean, and avian faunas from California, and were subsequently used by a number of others working in both the New World (e.g., Butler 2000; Dean 2001; Lyman 2003a, 2003b, 2004d) and the Old World (e.g., Butler 2001; Grayson and Delpech 1998; Nagaoka 2001, 2002; Stiner et al. 1999).

The sort of graph referred to is exemplified in Figure 5.15. This is a simple bivariate scatterplot of the abundance of North American elk (or wapiti, *Cervus elaphus*) remains relative to the abundance of all ungulate remains in eighty-six assemblages of mammalian remains from sites in eastern Washington State. In an earlier analysis, I found that the absolute abundance of elk remains is not correlated with the total ungulate sample size ( $\Sigma$ NISP) per assemblage ( $r = 0.16, p = 0.14$ ) (Lyman 2004d); removing the three largest collections (NISP > 1800), the correlation becomes weak but significant ( $r = 0.33, p < 0.005$ ). The relative abundance of elk per assemblage is not correlated with assemblage age ( $r = 0.006, p > 0.9$ ), and the slope of the simple best-fit regression line is not significantly different from zero. On this basis I suggested that there is no evidence here that the abundance of elk relative to the abundance of all ungulates changed over the 10,000 years represented (Lyman 2004d).

Since the earlier analysis, the  $\chi^2$  test for linear trends has been introduced to paleozoology. That test suggests there is indeed a trend in the relative abundance of elk remains over time ( $\chi^2$  trend = 112.96,  $p < 0.0001$ ). The scatterplot in Figure 5.15 gives no indication of whether the trend is for elk remains to increase or decrease over time. Thus one advantage of the sort of graph shown in Figure 5.15 is exemplified by the included simple best-fit regression line, which hints at a decrease in elk abundance over time (despite it having a slope statistically indistinguishable from no slope). Remember, however, that taxonomic abundance data are often at best only ordinal scale data, and seldom are they ratio scale data. Therefore, the regression line in Figure 5.15 should not be interpreted literally as indicative of ratio scale relative abundances of elk. Instead, that line assists with the identification of a trend in (in this case) elk relative abundances. Consider another example, one that displays a graphically visible trend.

The Emeryville Shellmound site on the shore of San Francisco Bay in California produced a large fauna from stratified deposits. Paleoziologist Jack Broughton

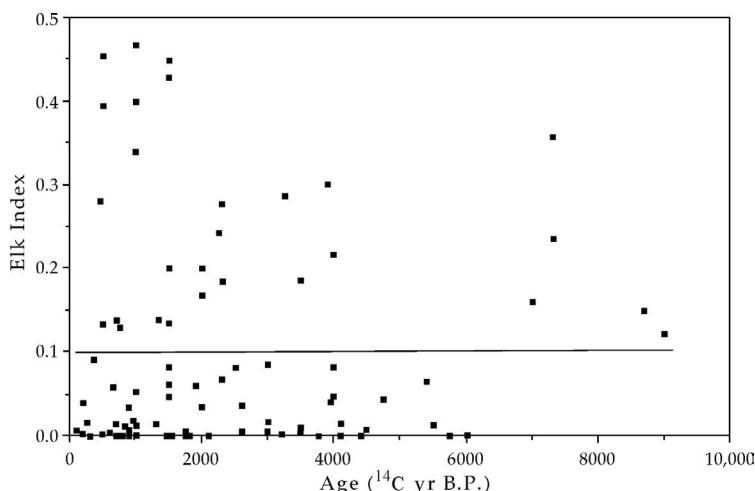


FIGURE 5.15. Bivariate scatterplot of elk abundances relative to the sum of all ungulate remains in eighty-six assemblages from eastern Washington State. The simple best-fit regression line is shown to assist with identifying trends in elk relative abundance.

(1999) analyzed those faunal remains and found a number of trends in taxonomic abundances. One of the more interesting ones concerns the abundance of remains of North American elk relative to the abundance of deer (*Odocoileus* sp.) remains. To monitor change in the abundance of elk over time, Broughton calculated an “elk–deer index” for each assemblage in each stratum (several strata produced more than one assemblage). The index was calculated as  $\sum \text{elk NISP} / (\sum (\text{elk NISP} + \text{deer NISP} + \text{medium artiodactyl NISP}))$ , where it is very likely that all medium artiodactyl remains are from deer. The data and index value for each assemblage are given in Table 5.9. As shown in Figure 5.16, plotting the index value, which is effectively the proportion of cervid remains that represent elk, against the stratum from which it derives suggests that elk became less available to human occupants of the Emeryville Shellmound site over time (stratum 1 is youngest, stratum 10 is oldest). That it is indeed the case that elk availability decreased over time relative to deer is confirmed by the simple best-fit regression line through the point scatter ( $r = 0.62$ ,  $p = 0.008$ ). The  $\chi^2$  test for trends also suggests that there is a trend in the relative abundance of elk remains ( $\chi^2$  trend = 638.8,  $p < 0.0001$ ), but does not indicate the direction of change in abundance.

Changes in taxonomic abundances can be monitored more directly than with the indices of relative elk abundances presented in Figure 5.16 as proportions of some larger group of taxa. Table 5.10 lists the abundances of remains of two taxa of

Table 5.9. Frequencies (NISP) of elk, deer, and medium artiodactyl remains per stratum at Emeryville Shellmound. Data from Broughton (1999)

Stratum	Elk	Deer	Medium Artiodactyl	Elk-Deer Index
1	0	100	122	0.0
1	0	35	58	0.0
2	8	758	958	0.005
3	17	294	365	0.025
3	1	8	12	0.048
4	81	72	76	0.354
5	40	52	56	0.270
6	11	20	18	0.224
7	12	10	13	0.343
7	184	94	76	0.520
8	75	46	56	0.424
8	116	87	83	0.406
9	23	79	51	0.150
9	68	67	62	0.345
10	50	95	94	0.209
10	59	122	103	0.208
10	63	105	88	0.246

mammals recovered from the stratified deposits of Homestead Cave in Utah State (Grayson 2000). The deposits span the terminal Pleistocene and entire Holocene. Based on pollen and plant macrofossil records, local climate shifted from cool and moist relative to today, to more or less modern climate by about 8,000 years ago. The two taxa were chosen from the several dozen represented in the site to illustrate trends in abundance here; *Neotoma cinerea* prefers cool, moist conditions relative to *Dipodomys* sp., which prefers warmer and drier conditions. The absolute abundance of *Neotoma cinerea* is not correlated with the stratum total NISP ( $\rho = 0.13, p > 0.6$ ); the absolute abundance of *Dipodomys* sp. is correlated with the stratum total NISP ( $\rho = 0.82, p = 0.0003$ ), but given the exceptionally large abundances of remains in all strata, I doubt that sample size is a problem.

Given the climatic preferences of the two taxa, and the environmental history coincident with deposition of the strata in Homestead Cave, *Neotoma cinerea* should decrease in abundance and *Dipodomys* sp. should increase. That is in fact precisely

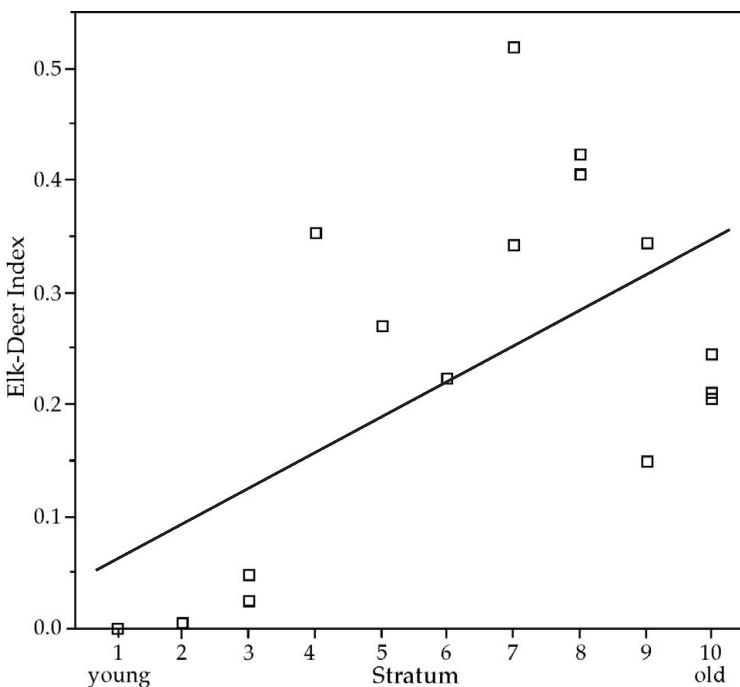


FIGURE 5.16. Bivariate scatterplot of elk–deer index against stratum at Emeryville Shellmound. The simple best-fit regression line is shown to assist with identifying trends in elk relative abundance. Data from Table 5.9.

what the relative abundances of remains of each taxon do. On the one hand, *Neotoma cinerea* remains decrease in relative abundance during the terminal Pleistocene and earliest Holocene (strata I, II, III) and virtually disappear after that (Figure 5.17;  $\chi^2$  trend = 13,034,  $p < 0.0001$ ). Remains of *Dipodomys* sp., on the other hand, increase rapidly during the terminal Pleistocene and earliest Holocene, and after about 8,000 BP (after the deposition of stratum VI), they comprise more than half the total NISP per stratum (Figure 5.17;  $\chi^2$  trend = 25,457,  $p < 0.0001$ ).

Two methods for examining trends in taxonomic abundances over time have been described. One involves calculating an index of a taxon's abundance within a limited set of taxa. The index can be expressed as a proportion or a percentage. The other method involves calculating the proportion or percentage of a taxon's abundance within the entire assemblage. Broughton (1999) was interested in determining whether elk availability was decreased by human predation and were replaced by deer (Figure 5.16). Grayson (2000) was interested in the contribution of particular taxa

Table 5.10. *Frequencies (NISP) of two taxa of small mammal per stratum at Homestead Cave. Remains from strata X, XIII, XIV, XV were not analyzed. Data from Grayson (2000)*

Stratum	<i>Neotoma cinerea</i>	Percent of Total NISP	<i>Dipodomys sp.</i>	Percent of Total NISP	Stratum Total NISP
I	2,577	25.1	360	3.5	10,275
II	1,508	19.2	1,329	16.9	7,855
III	306	10.6	1,056	36.6	2,884
IV	242	0.9	13,712	51.5	26,615
V	1	0.02	2,965	58.0	5,109
VI	5	0.02	15,173	62.4	24,330
VII	4	0.03	9,868	71.0	13,905
VIII	5	0.06	5,742	69.3	8,289
IX	1	0.005	15,477	70.1	22,088
XI	0	0	6,820	67.6	10,096
XII	0	0	16,753	73.3	22,860
XVI	0	0	4,371	69.4	6,296
XVII	9	0.06	10,418	67.0	15,548
XVIII	0	0	720	68.8	1,047

to the entire mammalian fauna. The choice of method was guided by the research question being asked.

## CONCLUSION

There are many reasons to compare the taxonomic abundances displayed by different collections. Simplistically, these can be reduced to two general categories of questions – those concerning paleoenvironmental conditions (was it hot or cool, dry or moist?), and those concerning human or hominid adaptations (what did they eat, and how much)? Those categories concern ultimate questions; they are answered with detailed contextual, associational, and taphonomic analysis of taxonomic abundances. The concern of this chapter has been to describe ways that taxonomic abundance data might be analyzed and studied, to answer more proximal questions, questions closer to the quantitative data themselves. Toward that end, *diversity* was defined as a generic term for variation in taxonomic richness, evenness, and/or heterogeneity. Indices for each of the latter variables were described and

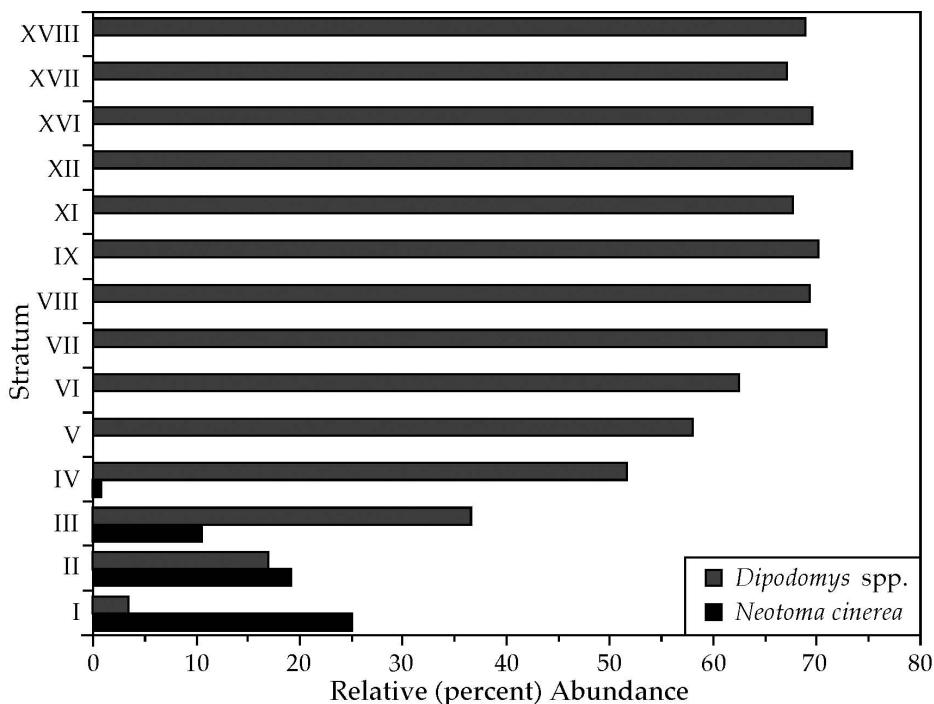


FIGURE 5.17. Relative abundances of *Neotoma cinerea* and *Dipodomys* spp. at Homestead Cave. Data from Table 5.10.

exemplified. Two methods of monitoring trends in taxonomic abundances over time were discussed. No doubt, there are other methods and indices that might be used. Ecologists are regularly inventing new measures of taxonomic structure and composition, and refining old ones. Paleozoologists should, and often do, pay attention to those developments and adopt new indices and quantitative methods developed by ecologists. In so doing, paleozoologists potentially adopt a family of problems the identification of which is a good way to conclude this chapter.

Earlier the biological concept of community was defined, and it was noted that biological communities are sometimes difficult to identify; the identification problem is exacerbated when one seeks to identify a paleocommunity on the basis of the fossil record (e.g., Bennington and Bambach 1996). This fact was explicitly dealt with more than 50 years ago by paleozoologist J. Arnold Shotwell (1955, 1958) when he attempted to use quantitative measures of skeletal completeness to distinguish taxa comprising the proximal paleocommunity from taxa comprising distal (distant) communities. Taphonomists and those with a quantitative bent were quick to point out some of the analytical difficulties with what Shotwell proposed (Grayson 1978b and references

therein). With the benefit of another couple decades of reflection, there is yet another problem that attends Shotwell's method, a problem that potentially plagues any and all of the indices and measures of taxonomic abundances discussed in this chapter. That problem is what is known as "time averaging."

The temporal resolution available in the paleozoological record is seldom of the fine scale, intrageneration resolution that is provided to zoologists. The temporal acuity of the paleozoological record is such that, very often, any stratigraphically bounded sample is a palimpsest or time-averaged collection representing multiple generations, multiple seasons, multiple years, and typically multiple decades or even centuries or millennia (Peterson 1977; Schindel 1980). This means that concepts such as taxonomic richness, evenness, and heterogeneity, which derive from extant communities, may never be of the same temporal resolution in the paleozoological record as they are in the modern or extant zoological record (but see Olszewski and Kidwell 2007). Ecological time is seldom the same as paleozoological time (see also Grayson and Delpech 1998). One result of recognition of this potential problem has been a series of "fidelity studies," introduced in Chapter 2 and defined as "the quantitative faithfulness of the [fossil] record of morphs, age classes, species richness, species abundance, trophic structure, etc. to the original biological signals" (Behrensmeyer et al. 2000:120). Many of these studies originate in taphonomic concerns regarding the preservation of animal remains or the rate of input of those remains. Fewer concern the difference between ecological time and paleontological time either studied empirically or focusing on key quantitative concepts such as richness and evenness (see Broughton and Grayson [1993], Lyman [2003b], and Olszewski and Kidwell [2007] for exceptions). The critical point to contemplate is the relationship between the property (richness, relative abundance, etc.) of a paleofauna that has been measured and the temporal resolution of that value. Does it encompass a decade, a century, more? And how might that influence interpretations? An example will highlight the significance of these questions.

If the elk abundance data for the eighty-six individual collections in Figure 5.15 are lumped into 500-year bins (1–500 BP, 501–1000 BP, etc.), and the elk index recalculated using only elk and deer remains (bison, pronghorn, bighorn excluded), the result is rather different than that shown in Figure 5.15. The resulting graph and best-fit regression line suggest there is a significant relationship between age and the relative abundance of elk ( $r = 0.489$ ,  $p = 0.064$ ) (Figure 5.18). The slope of the line suggests elk availability decreased over time. The  $\chi^2$  test for trends confirms the trend in the relative abundance of elk remains ( $\chi^2$  trend = 9.67,  $p = 0.002$ ; including all ungulates,  $\chi^2$  trend = 42.5,  $p < 0.0001$ ). Something not apparent in the statistics is that elk seem least abundant between about 7,500 and 4,000 BP, precisely

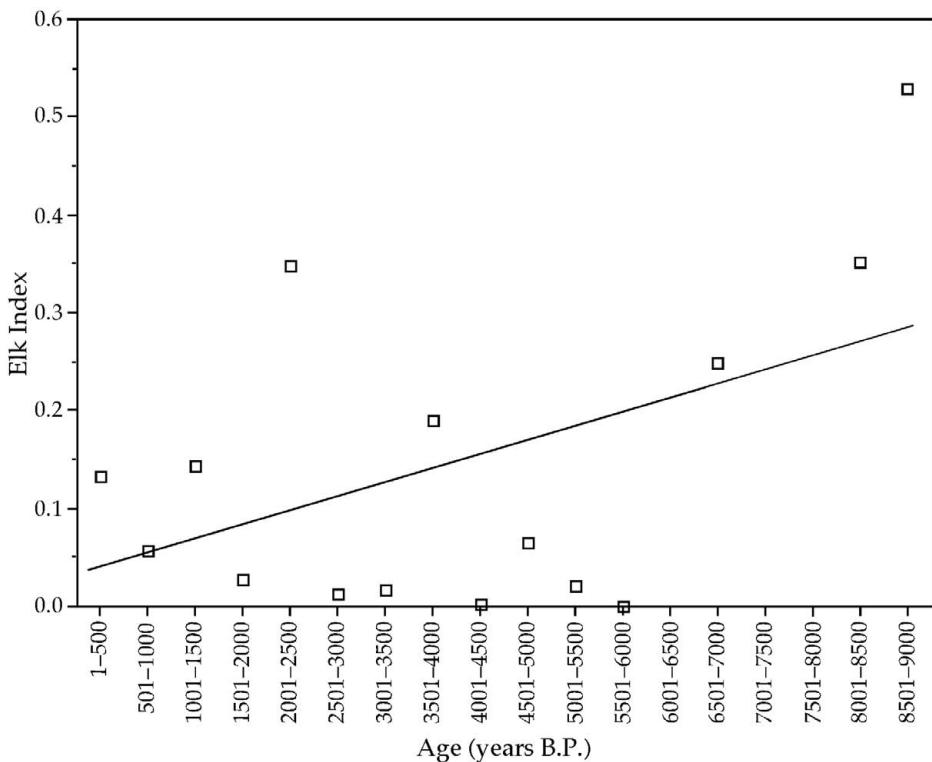


FIGURE 5.18. Bivariate scatterplot of elk abundances relative to the sum of all ungulate remains in eighty-six assemblages from eastern Washington State summed by age for consecutive 500-year bins. The simple best-fit regression line is shown to assist with identifying trends in elk relative abundance. Compare with Figure 5.15. Note that there are no data for the 6,001–6,500, 7,001–7,500, and 7,501–8,000 bins.

when climate was least conducive to elk reproduction (Lyman 2004a, 2004d, and references therein). Whatever the case, comparison of Figures 5.15 and 5.18 make it clear that time averaging can influence analytical results, as can the taxa included in the measure of relative abundance.

Do not let the preceding mislead you. Time averaging may not always be a bad thing. It has been suggested, for example, that either time averaged samples formed by natural processes, or those formed by analytical lumping of assemblages may serve to filter out “noise” in a paleofaunal signal (Olszewski 1999; see also Muir and Driver 2002). Whether time averaging is a good thing or a bad thing will depend on the research question being asked and the attendant target variable that must be measured or estimated in order to answer the research question. Similarly, lumping assemblages from different spatial locations may also result in an averaging or muting

of the paleofaunal signal (Lyman 2003b). Thus, one must be explicit about the spatiotemporal coordinates (and their boundaries) pertinent to the research question being asked. Either that, or we rewrite the ecological variables we seek to measure in paleozoological spatiotemporal units. To reiterate, the research questions of interest should dictate the temporal resolution necessary for a clear answer.

## Skeletal Completeness, Frequencies of Skeletal Parts, and Fragmentation

The minimum number of individuals (MNI) is typically defined as something like the most frequently occurring skeletal part (Table 2.4). Variation in how that definition can be operationalized are differences in size, age, sex, or recovery context (aggregation) considered renders MNI as a derived measure. But on a general, in some ways less discriminating scale, how the basic definition of MNI is operationalized is simply this: Given all the remains of a taxon in a collection (how the spatiotemporal boundaries of that collection are defined need not concern us initially), redundant skeletal parts are each tallied as a single MNI. *Redundant skeletal parts* means that specimens overlap anatomically. Two left femora of deer overlap anatomically and are redundant with one another, just as are two upper right second molars, three right distal humeri, and four left innomates. In the order listed, the MNI values are 2, 2, 3, and 4. To reiterate, redundant that is, anatomically overlapping skeletal parts each represent a unique individual or a tally of one MNI.

How MNI is operationalized on a general scale by redundant skeletal parts forces us to recognize a previously unmentioned quantitative unit. That unit was not really distinctively named until 1982, but it had played an important role in paleozoology for decades prior to that time. Today that quantitative unit is known as the minimum number of elements, or MNE. This unit not only is, in fact, the basis of MNI, it is also the basis of another important quantitative unit that has at least two different names (MAU,%survivorship), and it is, finally, somewhat of a misnomer. The purposes of this chapter are to make all of these points clear and to illustrate if and how MNE might be analytically useful.

Another purpose of this chapter is to review the means to analytically determine, and interpret, other quantitative variables that paleozoologists sometimes measure. One concerns which parts or portions of the represented skeletons are present; are those parts represented in abundances that reveal something of the taphonomic history of the collection, such as accumulation or preservation? Another concerns the

degree of skeletal completeness; how complete, on average, is each of the multiple skeletons represented? The final variable that has been studied using MNE concerns bone fragmentation. Techniques for quantifying the degree of fragmentation evident in a collection of faunal remains are reviewed, and pertinent concepts defined. Tallying the frequencies of skeletal parts, measuring skeletal completeness, and measuring the degree of fragmentation all rest in one way or another on tallies of  $MNE_i$ , where  $i$  is a particular skeletal element, part, or portion. An historical overview of MNE sets the stage for subsequent discussions of how it and its related quantitative units have been and can be used. The overview provides the requisite background to consideration of how skeletal completeness is measured and how fragmentation is quantified.

#### HISTORY OF THE MNE QUANTITATIVE UNIT

When Chester Stock tallied up abundances of mammalian taxa represented by the Rancho La Brea remains (Chapter 1), he determined the minimum number of individuals. To determine MNI, he tallied redundant or anatomically overlapping skeletal parts; that is, he determined the MNE of each unique skeletal part whether first cervical vertebrae, left mandibles, right humeri, or left tibiae and used the largest number as his MNI. All subsequent paleozoologists who have derived an MNI from a collection have first determined the MNE for at least the most common skeletal parts if not all skeletal parts of a taxon, and then used the greatest MNE value as the MNI value for that taxon. As put some years ago, MNE is the minimum number of skeletal portions necessary to account for the specimens representing that portion (Lyman 1994c:102). How and why did explicit recognition of MNE emerge?

Traditionally, paleozoologists such as Chester Stock sought measures of taxonomic abundances. Thus, MNI, biomass, NISP, and the like were designed to measure those abundances. But with increases in our knowledge of how the paleozoological record formed, taphonomic concerns came to the fore. Was, perhaps, taxon A more abundant than taxon B in a collection because taxon A was more abundant on the landscape than taxon B was at the time the remains of each were accumulated and deposited? Or, did the agent or process of bone accumulation simply collect more of taxon A than of taxon B? Or, did the agent or process of bone preservation (or destruction) result in bones of taxon A being more frequently preserved than those of taxon B? These were taphonomic questions that concerned skeletal-part abundances and, thus, depending on their answers, could significantly influence how taxonomic abundance data were interpreted.

Taphonomists seek answers to questions that differ in kind and scale from those traditionally asked by paleozoologists. (This is not meant to imply that paleozoologists and taphonomists are two distinct sets of researchers; rather, it is meant to underscore differences in the questions that are asked about a collection of faunal remains.) Instead of asking how many or how much of each taxon contributed to a collection what a paleozoologist would normally ask, although perhaps with different target variables in mind (Figure 2.1) a taphonomist would ask why only certain skeletal elements, individual organisms, or taxa are represented in a collection. Taphonomists attempt to measure and understand why paleozoologists do not always find complete articulated skeletons. Rather, what they typically find are variously incomplete skeletons, often comprising broken skeletal elements. The questions taphonomists ask, then, are focused more on the proximate causes for the existence of a bone assemblage and why the assemblage is made up not only of the taxa that it represents, but the skeletal parts and the taxonomic abundances that it is, as well as such questions as why some bones are missing, others are broken, and still others are abundant.

Taphonomists effectively shift the level of measurement from taxonomic abundances to attributes of a collection of remains, such as skeletal-part frequencies, usually of a single taxon at a time (Lyman 1994c). Paleozoologists with taphonomic interests have an advantage over other scientists who study the past such as paleobotanists; the former have a model of what a complete animal carcass consisted of skeletally. If it was a mammal, then it had one skull, left and right mandibles, (most likely) thirteen left and thirteen right ribs, thirteen thoracic vertebrae, left and right humeri, and so on. A paleobotanist has no such model of, say, an apple tree; the questions How many leaves? How many apples? How many seeds? cannot be answered with any confidence for want of a model of a standard apple tree with  $N_1$  leaves,  $N_2$  apples, and  $N_3$  seeds. The model of a complete animal is where a paleozoological taphonomist starts. Measuring how a collection of bones and teeth of a taxon differ from a collection of complete skeletons of that taxon is the first step toward answering the ultimate taphonomic question: How and why are *these* bones and teeth *here* and in the condition that they are whereas other bones and teeth are missing or in different condition (Lyman 1994c)?

Increasing demand to answer more taphonomically detailed questions in order to build strongly warranted explanations of the paleozoological record in the 1960s and 1970s brought about a growth in the number of quantitative units used to measure aspects of the record (Lyman 1994a). Perhaps not surprisingly, then, the first explicit use of MNE is found in a seminal, explicitly taphonomic study. Voorhies (1969) sought to understand the taphonomy of an early Pliocene paleontological site in

the state of Nebraska. He did not use the term MNE but listed what he termed the number of individual skeletal elements represented for twenty-six different skeletal elements of an extinct form of antelope the remains of which made up the bulk of the collection he studied. Voorhies (1969:1718) described the quantitative unit as the minimum number of elements (bones) represented by all identifiable fragments of the element in the collection, and distinguished it from the [minimum] number of individual *animals* represented by nonserial paired [skeletal] elements.

Voorhies designed the MNE quantitative unit because he was concerned with why some skeletal parts were very frequent in the collection whereas others were of mid-level abundance and still others were rare. Observed MNE frequencies of skeletal parts diverged from the model of a set of complete skeletons, and Voorhies wanted to know why that divergence existed. He began by trying to determine if a pattern in skeletal-part frequencies existed, and to do that required that he use an MNE-type quantitative unit. He sought an answer to a question concerning taphonomy, in particular a question concerning skeletal-part abundances. He found that skeletal elements that were of light weight relative to their volume tended to be winnowed out of a deposit by flowing water whereas bones that were heavy relative to their volume lagged behind (were not removed by flowing water). His research resulted in the designation of what came to be known as Voorhies Groups distinctive groups of particular skeletal elements that are variously winnowed, moved, or left behind by the action of flowing water (e.g., Behrensmeyer 1975; Lyman 1994c; Wolff 1973).

About a decade after Voorhies (1969) used the MNE quantitative unit, zooarchaeologists seem to have independently invented the same unit. The question they asked was ultimately the same one asked by Voorhies: Why were some skeletal parts abundant and other parts rare in a collection? Binford (1978, 1981, 1984; Binford and Bertram 1977) was interested in how hominids differentially dismembered and transported portions of prey carcasses, and wanted to identify the effects of natural attrition such as carnivore gnawing on frequencies of skeletal parts. Thus, Binford, like Voorhies, started with the model of a complete carcass or skeleton (or multiple complete carcasses or skeletons), and sought to explain why collections of prehistoric bones had, say, more thoracic vertebra than ribs, or more distal humeri than proximal humeri. Binford initially defined the quantitative unit he designed as the minimum number of individual animals represented by each anatomical part and referred to them as MNI values (Binford and Bertram 1977:79). Binford (1984:50) made it clear that his MNI values were not the same as the traditional MNI values of Stock (1929), White (1953a), and Chaplin (1971), when he stated that I have decided to reduce the ambiguity of language by no longer referring to anatomical frequency counts as MNIs.

Bunn (1982) is the first paleozoologist I know of to use the term MNE. Bunn (1982:35) defined the unit as the minimum number of elements, but he did not define element. He, like Binford (1978, 1981; Binford and Bertram 1977) before him, used MNE to signify not only anatomically complete skeletal elements (e.g., femora), but anatomically incomplete skeletal parts (e.g., the distal half of humeri) and also portions of a skeleton made up of multiple discrete skeletal elements (e.g., thoracic portion of vertebral column). This makes MNE even more of a derived unit than traditional MNI values. Both MNE and MNI might be determined with or without consideration of age, sex, and size differences among specimens (is a particular fragment of a proximal left humerus from a small female, 2-year-old deer, distinguished from another fragment of a proximal left humerus from a large male, 5-year-old deer that does not anatomically overlap the first?). But whereas an individual animal is a natural, inherently bounded unit (its boundaries are its skin), a femur is a skeletal element that is naturally bounded (it is discrete), but a distal femur is not naturally bounded and a thoracic section of the vertebral column is not necessarily discrete. These sorts of considerations influence MNE values, and there are even more fundamental issues to contend with when it comes to deciding if femora outnumber humeri and the like. There are other as yet unmentioned phases to the history of MNE, but discussion of them is better served if they come up later. The next issue that must be dealt with is how MNE values are determined.

#### DETERMINATION OF MNE VALUES

Despite a relatively simple definition of MNE as the minimum number of skeletal elements necessary to account for the specimens under study, this quantitative unit has seen more discussion and debate over how it is to be determined during the past 20 years or so than any one might have imagined. This is because MNE and two quantitative units based on it (not including MNI) became extremely important to answering taphonomic questions about skeletal-part abundances beginning in the 1970s (e.g., Andrews 1990; Binford 1981, 1984; Dodson and Wexlar 1979; Hoffman 1988; Klein 1989; Korth 1979; Kusmer 1990; Lyman 1984, 1985, 1994b, 1994c; Marean and Spencer 1991; Shipman and Walker 1980; Turner 1989).

Many researchers suggested ways to determine MNE. For example, Klein and Cruz-Uribe (1984:108) proposed that each specimen be recorded as a fraction using common and intuitively obvious fractions (e.g., 0.25, 0.33, 0.5, 0.67) and not attempting great precision. The fractions were used to record how much of a category of skeletal part (typically a proximal or distal half of a long bone) each specimen rep-

Table 6.1. *MNE values for six major limb bones of ungulates from the FLK Zinjanthropus site*

Skeletal element	Bunn (1986); Bunn and Kroll (1988)	Potts (1988)
Humerus	20	19
Radius	22	18
Metacarpal	16	14
Femur	22	8
Tibia	31	12
Metatarsal	16	16

resented; all recorded fractions were summed to estimate the MNE for a category of skeletal part. Thus, if the category is proximal half of the femur, the analyst records a specimen representing the proximal half of a femur as 1.0, a specimen representing one half of a proximal femur is recorded as 0.5, and a specimen representing one third of a proximal femur as 0.33. Adding those fractions produces an MNE of 1.83, or an MNE of two proximal femora halves. This procedure does not, however, take account of anatomical overlap. For example, what if the three specimens of proximal femur all include the greater trochanter? If they do, the MNE is not two, but three.

Marean and Spencer (1991:649650) suggested measuring the percent of the complete circumference represented by long-bone diaphysis specimens. Summing the percentages recorded for each portion of the diaphysis measured say, proximal diaphysis, middiaphysis, and distal diaphysis would provide an estimate of the MNE of particular shaft portions. Again, the weakness is that anatomically overlapping skeletal parts are ignored, potentially resulting in under estimates of MNE. Bunn and Kroll (1986, 1988) described three ways to derive MNE values from a collection of mammalian long bones. The analyst may determine (1) the minimum number of anatomically complete limb skeletal elements necessary to account for only the specimens with one or both articular ends, (2) the minimum number of complete limb skeletal elements necessary to account for only the specimens of limb diaphyses (without an articular end), and (3) the minimum number of complete limb skeletal elements necessary to account for both the specimens with one or both articular ends and the diaphysis specimens lacking articular ends. These are labeled MNEends, MNEshafts, and MNEcomp, respectively.

Underscoring that the quantitative unit is derived, if the method of determining MNE varies, different MNE tallies are produced. Table 6.1 presents MNE values for the six major limb bones of ungulates represented in the Plio-Pleistocene FLK

*Zinjanthropus* collection. MNE values were calculated by two researchers who used different methods. Bunn (1991) used all specimens diaphysis fragments as well as epiphyses whereas Potts (1988) focused mostly on articular ends. Given what we have seen in preceding chapters and the consistently strong relationship between NISP and MNI, and between sample size (however measured) and a variable of interest, it should come as no surprise that if Bunn included more long-bone diaphysis specimens in his derivation than did Potts, then Bunn's MNE values would be greater than Pott's.

MNE is defined in precisely the same way that MNI is defined but at the scale of a partial skeleton (what I have been referring to as a skeletal part or portion) rather than at the scale of a complete skeleton (Lyman 1994b). To reiterate using different wording, MNE is defined as the minimum number of skeletal elements necessary to account for an assemblage of specimens of a particular skeletal element or part (discrete item) or portion (multiple discrete items, such as all thoracic vertebrae in a vertebral column) (Lyman 1994b:289). The definition is operationalized by examining specimens of each kind of skeletal element or part, say, left femora or distal right humeri, for anatomical overlap (e.g., Bunn 1986; Morlan 1994). If three specimens of left femora comprise the collection (assuming all are of the same taxon), then the possible MNE represented by those specimens ranges from 1 to 3. If two specimens represent the complete distal end and one represents the proximal end, then the MNE of left femora is two. Just as when two left scapula and one right scapula are said to represent a minimum of two individuals (MNI = 2), anatomically overlapping skeletal specimens must represent unique individuals, whether these are individual organisms or individual skeletal elements.

The techniques used to determine MNE have, in the past 15 years, become hotly contested issues in paleozoology. Bunn (1991) and Potts (1988) used different methods to determine MNE values; the difference was whether or not diaphysis fragments were included in the tallies or only articular ends of long bones. That difference alone spawned a huge debate, in the literature, on whether a tally of skeletal-part frequencies was valid if diaphysis fragments were not included (e.g., Bartram and Marean 1999; Marean and Frey 1997; Pickering et al. 2003; Stiner 2002). A recent effort to bring all concerned parties together did not resolve the debate, though concessions were made (e.g., Cleghorn and Marean 2004; Marean et al. 2004; Stiner 2004). The debate originated in part with assumptions by some commentators regarding how other researchers were thought to count skeletal-part frequencies. In particular, those who advocate tallying diaphysis fragments have assumed that many have not counted diaphysis fragments but instead tallied only taxonomically diagnostic articular ends of long bones. This assumption is suspect given that ribs and many vertebrae are not taxonomically diagnostic, yet these specimens were tallied by those who allegedly did

not tally nontaxonomically diagnostic long-bone diaphysis fragments. Whatever the case, the important point is this: We must be explicit about how we count, whether we count NISP, MNE, MNI, or any other measure. Many paleozoologists still do not distinguish skeletal specimens, elements, parts, fragments, and the like, despite numerous suggestions over the past several decades that they be distinguished via explicit definitions (e.g., Casteel and Grayson 1977; Grayson 1984; Lyman 1994a, 1994b).

Assuming that all specimens in a collection, whether of articular ends or diaphysis fragments, are included in MNE (and NISP) tallies, we are left with the question of how to tally those specimens in order to derive an MNE value. The practical aspects of operationalizing the definition of MNE as based on anatomically overlapping specimens became technologically more sophisticated when Marean et al. (2001) applied GIS image software to the problem. They had found verbal descriptions and individual drawings of specimens that were anatomically incomplete skeletal elements to be cumbersome to manipulate when determining MNE values. For them, computer software provided a solution. Each specimen is outlined on a template of the skeletal element it represents; the template has previously been loaded into the software program. The software allows the analyst to digitally overlay outlines of multiple specimens. The maximum number of overlaps detected by the software indicates the MNE. The key step to the process, whether using hard-copy drafting paper or computer software, is drawing the specimens accurately. Marean et al. (2001) do not address this most critical step. Efforts by a student in my zooarchaeology class to replicate drawings of the same set of fragments of known (experimental) derivation found that the general shape of the fragment could be reproduced fairly consistently, but its size and location varied from iteration to iteration. Thus, some fragments that did not overlap in reality were sometimes drawn in such a way as to overlap in the database, and other fragments that did overlap in reality were sometimes drawn so as to not overlap. Errors increased in frequency and magnitude (degree of overlap, or lack thereof) as specimens displayed fewer and fewer anatomical landmarks. Tallies of MNE produced using the software varied from the actual number of elements by anywhere from 0 to more than 50 percent greater than the actual number across several trials.

The errors described in the preceding paragraph may be the result of inexperience, but the student who performed them earns her living applying GIS to archaeological problems, so degree of experience does not seem to be a significant factor. Additional trials by others using specimens of known derivation (e.g., experimentally generated from anatomically complete skeletal elements) might prove revealing, but I hazard the guess that the smaller the fragments one tries to draw the more errors in tallying MNE will result. Whether this is found to be true may be academic. This is so

because MNE is typically strongly correlated with NISP (Grayson and Frey 2004). This should not be surprising at all given that we know NISP and MNI are often strongly correlated, and we know that MNI is based on MNE. The relationship between MNE and NISP is examined in a subsequent section on fragmentation. Several other issues need to be dealt with first.

### *MNE Is Ordinal Scale at Best*

Given the definition of MNE as the number of skeletal parts or portions necessary to account for the specimens under study, it should be clear that MNE is but MNI at a less skeletally inclusive scale. Both are derived measures. Quite simply (and sadly) this means that all of the problems that attend MNI must also attend MNE. Furthermore, at best MNE can be only ordinal scale. To ensure these problems are appreciated, lets quickly review the major benefit and the major problem with MNI, but in terms of MNE.

MNI was designed to control for specimen interdependence when measuring taxonomic abundances. Thus, if two or more specimens came from the same individual, NISP would tally that individual twice but MNI would tally that individual only once. The same benefit attends MNE at the level of skeletal part or portion. Determination of MNE ensures that each skeletal part (or portion) that contributed to a collection will not be tallied, say, twice if represented by two specimens. This takes care of the treacherous problem of differential fragmentation that causes some paleozoologists to use MNI rather than NISP when measuring taxonomic abundances, and it takes care of the same problem at the level of skeletal part or portion when tallying abundances of ribs, cervical vertebrae, and tibiae. If humeri are broken into three pieces (on average) but femora are broken into six pieces (on average), the NISP of humeri will be half the value of the NISP of femora simply because of differential fragmentation and the attendant specimen interdependence. MNE values circumvent these problems. Varying degrees of interdependence of specimens representing a skeletal part or portion such as proximal (half of the) humerus, left half of the rib cage, or thoracic section of the vertebral column could influence relative NISP values of skeletal parts and portions. MNE might seem, therefore, to be a better unit than NISP for quantifying the abundances of skeletal parts and portions because it controls for specimen interdependence.

But, alas, MNE, like MNI, is subject to two serious problems. First, MNE is just that it is a *minimum*. Therefore, one cannot statistically compare two minimum values that might differentially range to some maximum value. Second, MNE is influenced

*Table 6.2. Fictional data showing how the distribution of specimens of two skeletal elements across different aggregates can influence MNE. Assuming no anatomical overlap of specimens, if stratigraphic boundaries are ignored, seven right and six left humeri, and fourteen right and nine left femora are tallied. If stratigraphic boundaries are used to define aggregates, nine right and eight left humeri, and fourteen right and ten left femora are tallied. R, right, L, left; P, proximal, D, distal*

	Humerus	Femur
Stratum 1	6 R P; 2 L P; 3 R D; 3 L D	4 R P; 1 L P; 4 L D
Stratum 2	1 R P; 4 L P; 1 R D; 1 L D	4 R P; 1 L P; 3 R D; 1 L D
Stratum 3	2 R D; 1 L D	6 R P; 5 L P; 4 R D; 4 L D

by sample size, aggregation, and definition (are sex, age, size taken into account). Different aggregates of specimens will often produce different MNE values, especially as sample sizes grow larger. If a collection is treated as one aggregate, the MNE of tibiae may be five, but if that collection is divided into three aggregates the MNE of tibia will likely increase because the parts that are redundant may be proximal ends in one aggregate, diaphyses in another, and distal ends in the third aggregate. Consider the data in Table 6.2, which is based on Table 2.13 where the influence of aggregation on MNI is illustrated. To produce Table 6.2 from Table 2.13, Taxon 1 in the latter was converted to humerus and Taxon 2 to femur, and humerus in Table 2.13 was converted to proximal in Table 6.2 and femur to distal. Assuming that there is no anatomical overlap of the skeletal parts in Table 6.2, the fictional data there show that exactly the same sorts of influence of aggregation can occur at the scale of skeletal element, part, or portion (MNE), as occur at the scale of individual animal (MNI). If stratigraphically defined aggregates are ignored and specimens are treated as one aggregate, there are seven right and six left humeri (= 13 humeri) and there are fourteen right and nine left femora (= 23 femora) listed in Table 6.2.

If the remains in each stratum in Table 6.2 are treated as comprising distinct aggregates, then the total number of humeri increases to nine right and eight left (= 17 humeri) and the total number of femora increases to fourteen right and ten left (= 24 femora). Note that if the remains are treated as one aggregate, the ratio of humeri to femora is 13:23 (or 0.565), but if the remains are treated as three separate aggregates, the ratio of total humeri to total femora is 17:24 (or 0.708). Just as with altering the aggregates of most common (redundant) skeletal specimens alters the resulting MNI,

Table 6.3. NISP and MNE per skeletal part of deer and wapiti at the Meier site

Skeletal part	Deer NISP	Deer MNE	Wapiti NISP	Wapiti MNE
Mandible	192	58	28	9
Atlas	44	22	3	2
Axis	19	17	5	5
Cervical	77	22	20	12
Thoracic	75	53	27	20
Lumbar	104	32	35	18
Rib	221	110	61	35
Innominate	130	43	34	15
Scapula	73	45	12	6
Humerus	150	58	26	11
Radius	164	60	40	15
Ulna	102	60	17	10
Metacarpal	133	87	34	10
Femur	86	29	34	13
Patella	13	11	2	2
Tibia	190	88	30	11
Astragalus	127	118	18	18
Calcaneum	159	121	18	16
Naviculo-cuboid	86	75	9	9
Metatarsal	143	89	48	12
First phalanx	224	148	86	58
Second phalanx	158	109	68	47
Third phalanx	75	75	25	25

changing the aggregates of most common (redundant) skeletal specimens alters the resulting MNE.

Sample size rendered as NISP also influences MNE values, as recently shown by Grayson and Frey (2004). Consider the deer (*Odocoileus* sp.) and wapiti (*Cervus elaphus*) data from the Meier site (Table 6.3). For deer, the NISP and MNE data are strongly correlated (Figure 6.1,  $r = 0.883$ ,  $p < 0.0001$ ), and the same holds for the wapiti data (Figure 6.2,  $r = 0.837$ ,  $p < 0.0001$ ). Grayson and Frey (2004) present numerous other examples in which the NISP per skeletal part and the MNE of skeletal parts are correlated. Their results and those for the Meier site deer and wapiti indicate that MNE values are often strongly influenced by sample size measured as NISP. The larger the NISP value per skeletal part or portion, the larger the MNE value for that part or portion. MNE is also influenced by how it is defined in the sense of whether

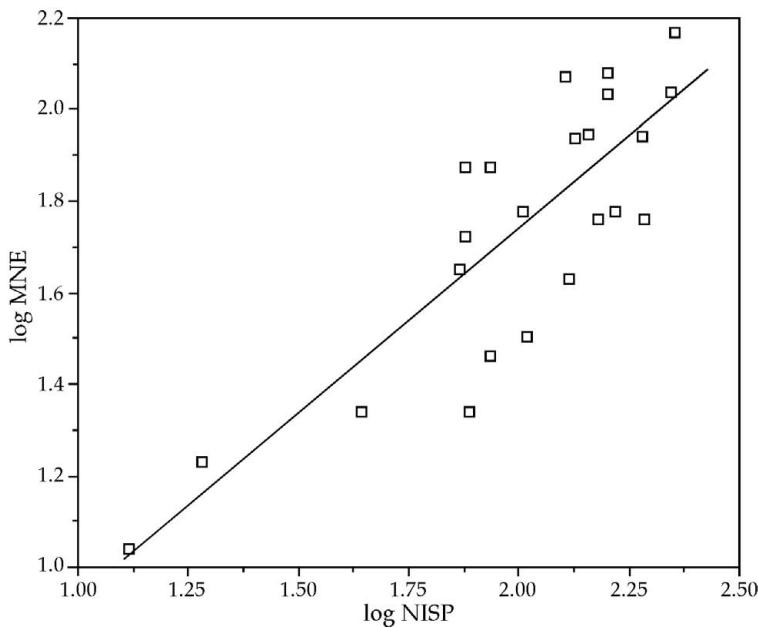


FIGURE 6.1. Relationship of NISP and MNE values for deer remains from the Meier site. Best-fit regression line ( $Y = 0.126X^{0.807}$ ;  $r = 0.837$ ) is significant ( $p = 0.0001$ ). Data from Table 6.3.

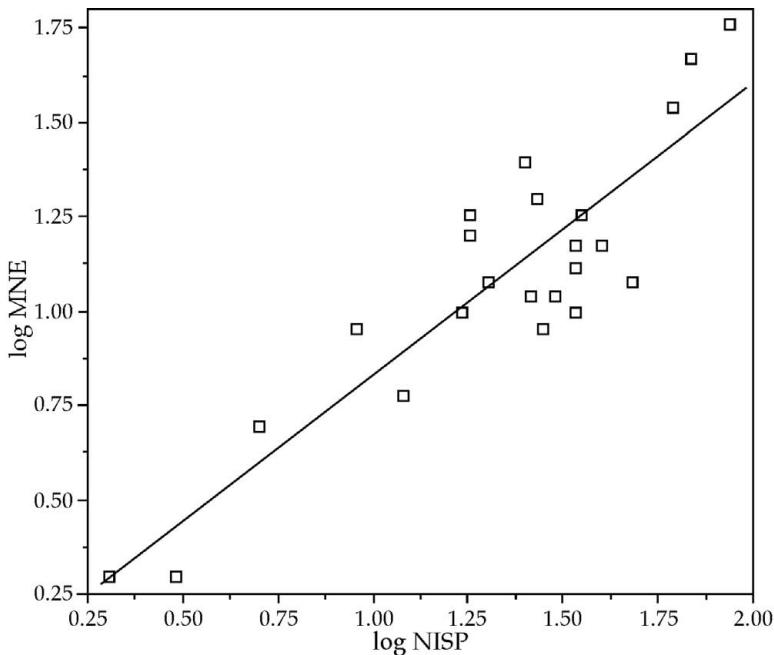


FIGURE 6.2. Relationship of NISP and MNE values for wapiti remains from the Meier site. Best-fit regression line ( $Y = 0.063X^{0.692}$ ;  $r = 0.883$ ) is significant ( $p = 0.0001$ ). Data from Table 6.3.

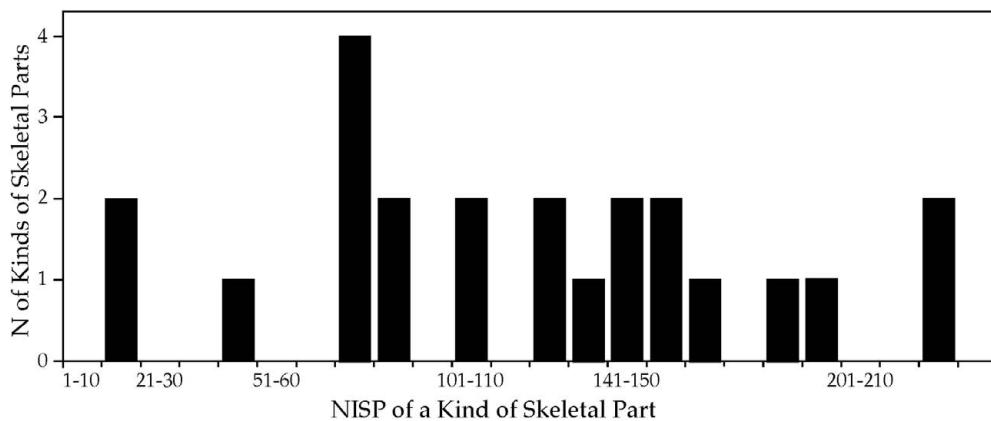
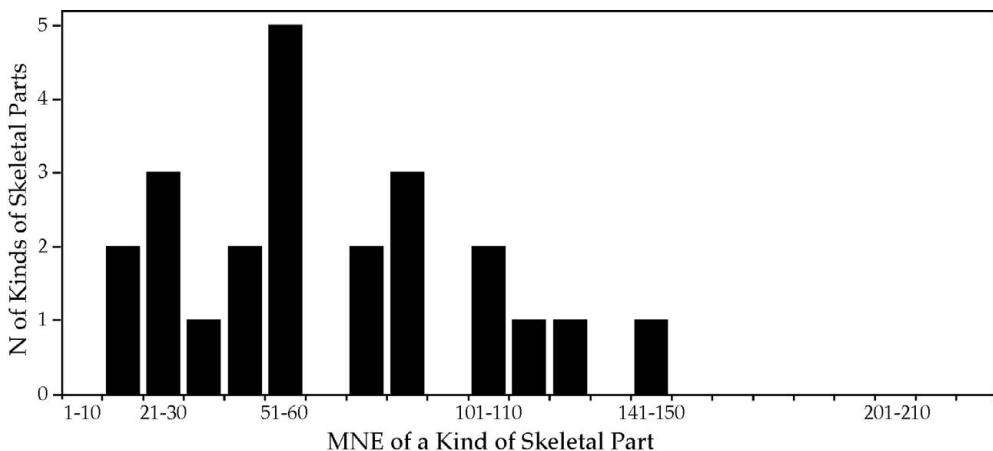


FIGURE 6.3. Frequency distributions of NISP and MNE abundances per skeletal part for deer remains from the Meier site. Data from Table 6.3.

or not size, ontogeny, and the like are taken into account when seeking to determine if two or more nonanatomically overlapping specimens come from the same original skeletal element.

Finally, MNE is at best ordinal scale. This can be shown using the same technique that was used to show that MNI is often at best ordinal scale (Chapter 2). The NISP and MNE data from Table 6.3 for the Meier site deer and wapiti are graphed in Figures 6.3 and 6.4, respectively. Note that the distributions of frequencies are different than those across taxonomic abundances illustrated in Figures 2.132.16. This is likely because in any given skeleton, there is a standard frequency of skeletal elements such that the frequency distribution is right skewed (like that observed for taxonomic abundances) but the mode is not to the farthest left but instead is slightly to the right (unlike that observed for taxonomic abundances)

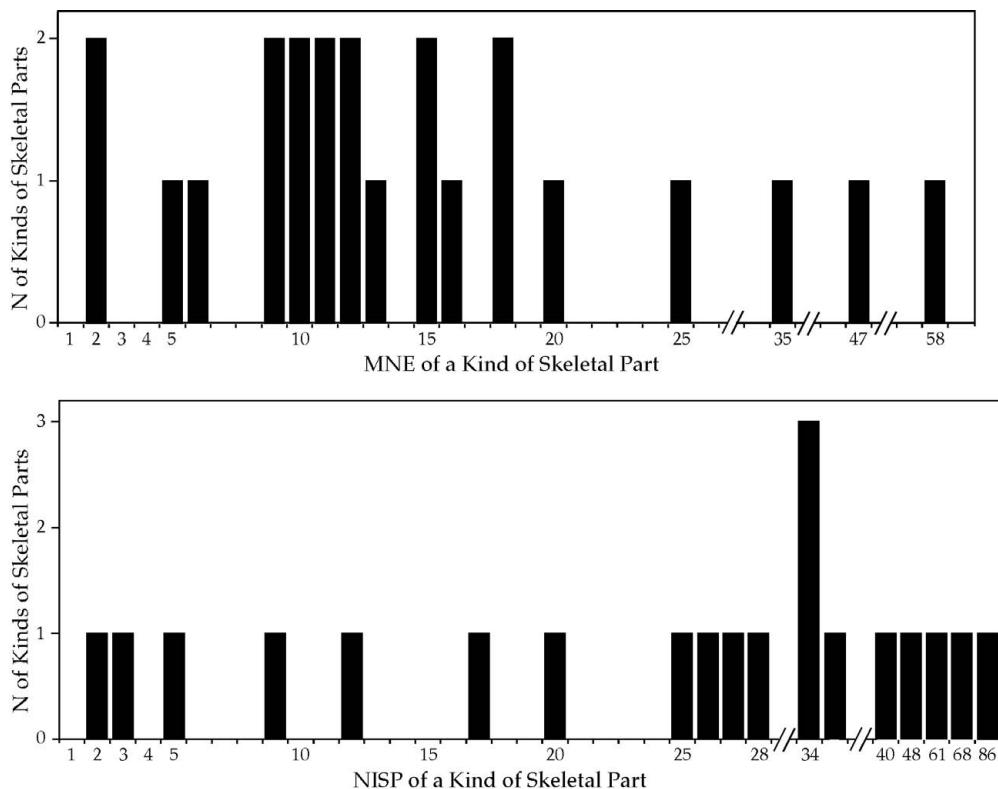


FIGURE 6.4. Frequency distributions of NISP and MNE abundances per skeletal part for wapiti remains from the Meier site. Data from Table 6.3.

(Table 6.4, Figure 6.5). Given that most paleozoological samples derive from  $> 1$  individuals—the taphonomic starting point (prior to variation in accumulation, preservation, and recovery) of a paleozoological collection—it is unlikely that a more or less random sample of remains from those multiple individuals will produce an extremely right-skewed frequency distribution. Rather, it is likely that few kinds of skeletal element will be represented by just one specimen or one MNE but instead most kinds of skeletal element will be represented by multiple specimens and MNE will be  $> 1$ .

Whatever the reason for the distributions observed in Figures 6.3 and 6.4, what is most important in the frequency distributions of NISP and MNE values of deer and wapiti at the Meier site is that there are gaps between many of the NISP values and particularly between the MNE values (Figures 6.3 and 6.4). Change in how the MNE values were defined may alter their ratio scale differences but is less likely to alter their ordinal scale rank order abundances. The same is likely for variation in aggregation and sample size. It is for these reasons that MNE values are ordinal scale. Their ratio

Table 6.4. *Frequencies of major skeletal elements in a single mature skeleton of several common mammalian taxa*

Skeletal element	Bovid/Cervid	Equid	Suid	Canid
cranium	1	1	1	1
mandible	2	2	2	2
atlas	1	1	1	1
axis	1	1	1	1
cervical	5	5	5	5
thoracic	13	18	14	13
lumbar	6 (or 7)	6	6 (or 7)	7
sacrum	1	1	1	1
innominate	2	2	2	2
rib	26	36	28	26
sternum	1	1	1	1
scapula	2	2	2	2
humerus	2	2	2	2
radius	2	2	2	2
ulna	2	2	2	2
carpal	12	14	16	14
metacarpal	2	2	8	10
femur	2	2	2	2
patella	2	2	2	2
tibia	2	2	2	2
fibula	2	2	2	2
astragalus	2	2	2	2
calcaneum	2	2	2	2
other tarsals	6	8	10	10
metatarsal	2	2	8	10
first phalanx	8	4	16	20
second phalanx	8	4	16	16
third phalanx	8	4	16	20

scale abundances likely will vary with aggregation, definition, and sample size, but their ordinal scale abundances likely will not, though this should be evaluated for each assemblage.

Much of the energy to develop and refine a protocol for determination of MNE values has been spent for little gain in mathematical or statistical resolution. Hours of refitting fragments that are otherwise unidentifiable would be better spent doing other things. Increasing the accuracy of drawing skeletal specimens whether by hand

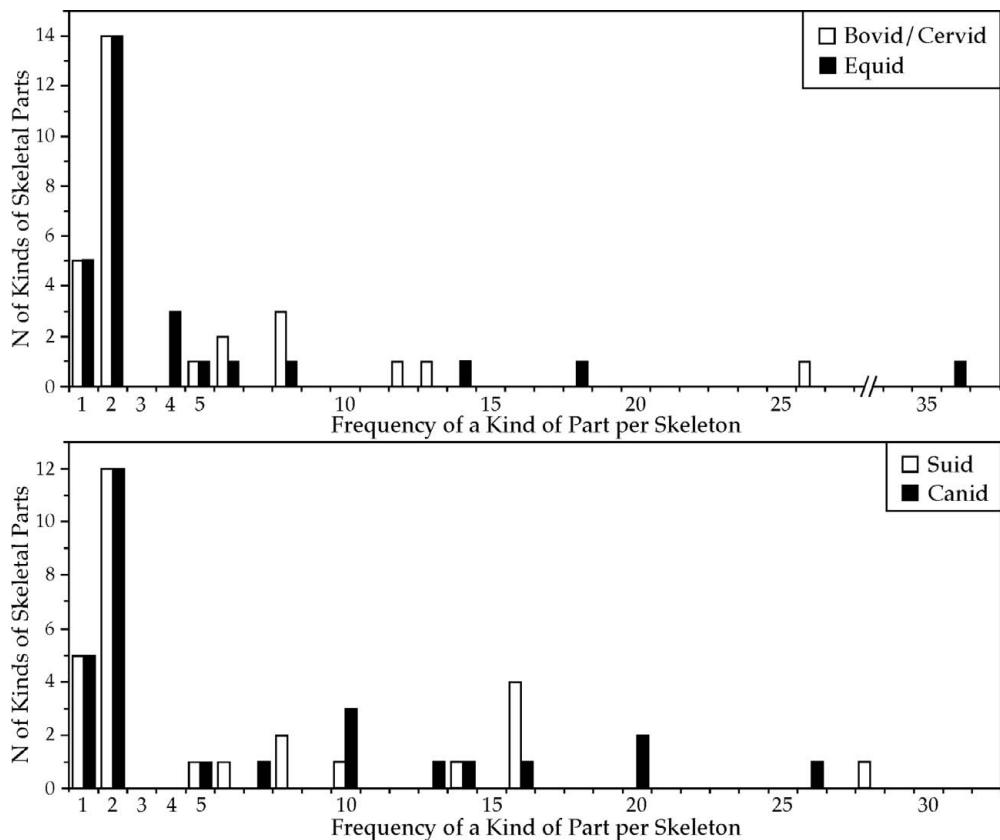


FIGURE 6.5. Frequency distribution of skeletal parts in single skeletons of four taxa.

or with the assistance of computer-aided sophisticated technology for purposes of detecting anatomical overlap is not a hoped for panacea.

#### *A Digression on Frequencies of Left and Right Elements*

Perhaps the simplest, as well as one of the earliest discussions of determining MNE values (without using the term MNE), was provided by White (1953a:397), who not only defined MNI as the most frequent of either left or right elements of a species, but also listed MNI values per skeletal part (e.g., proximal humerus, distal tibia). Interestingly, he also sometimes listed the frequencies of both left and right skeletal parts (White 1952, 1953b, 1955, 1956). In the latter, he was using MNE values, and importantly, he suggested that to divide [the total MNE, or sum of lefts and

Table 6.5. MNE frequencies of left and right skeletal parts of pronghorn from site 39FA83. P, proximal; D, distal. Original data from White (1952)

Skeletal part	Left	Right
Mandible	18	19
Innominates	13	19
Scapula	24	24
P humerus	3	0
D humerus	26	30
P radius	28	25
D radius	23	23
P ulna	23	22
P metacarpal	27	11
P femur	11	6
D femur	6	10
P tibia	9	9
D tibia	19	31
P metatarsal	22	15

rights per paired element or portion] by two would introduce great error because of the possible differential distribution of the kill (White 1953a:397). He, like Voorhies and Binford some years later, was interested in taphonomic questions regarding frequencies of skeletal parts, and he observed that in most of the features in the sites from which I have identified the bone the discrepancy between the right and left elements of the limb bones was too great to be accounted for by accident of preservation or sampling.... One should look for large discrepancies between the [frequencies of] right and left elements. Small discrepancies are not necessarily significant because they might be due to the accidents of sampling or preservation (White 1953b:59, 61).

White believed that human hunters, butchers, and consumers of animals might distinguish between the left and right sides of an animal, and butcher, transport, distribute, or discard the two sides of a large mammal differentially, yet he did not attempt to find evidence for this in any of the bone assemblages he studied. No one has, in the 50 years since White suggested it, sought such patterns in the frequencies of bilaterally paired bones. It is easy to illustrate how this might be done using data White (1952) published. The data involve MNE frequencies for pronghorn (*Antilocapra americana*) from archaeological site 39FA83 in the state of South Dakota (Table 6.5). If the MNE values were equivalent for left and right elements, then the

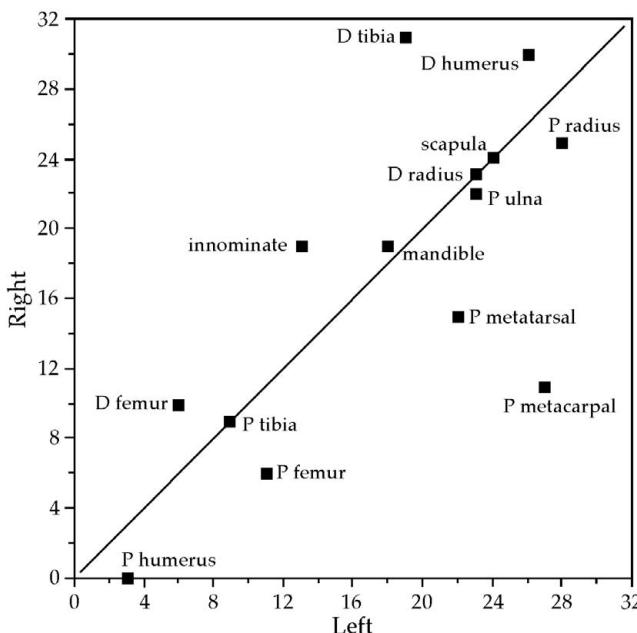


FIGURE 6.6. Comparison of MNE of left skeletal parts and MNE of right skeletal parts in a collection of pronghorn bones. Diagonal shown for reference. P, proximal; D, distal. Data from Table 6.5.

points plotted in Figure 6.6 should all fall on or near the diagonal line. Many of those points do not fall near the diagonal line. Do those points that do not fall on the line not fall there because of statistically significant differences between the frequencies of left and right specimens? Indeed, that seems to be the case for at least two, and perhaps three, skeletal elements. The adjusted residuals for each category of skeletal part indicate that there are more left proximal metacarpals (relative to few right proximal metacarpals) and more right distal tibiae (relative to few left distal tibiae) than chance alone would produce (Table 6.6; adjusted residuals are read as standard-normal deviates [Everitt 1977]). There may also be more left proximal humeri (and fewer right proximal humeri) than chance alone would produce, but the total small sample size of proximal humeri may be influencing the result.

Precisely this sort of analysis might be performed in order to accomplish what White suggested to determine if there is a significant difference in frequencies of left and frequencies of right elements. This could be important to assessing skeletal completeness, and also to evaluating, say, frequencies of proximal and distal halves of long bones. Both potentialities lead us to the next topic measuring the frequencies of particular skeletal parts.

Table 6.6. *Expected MNE frequencies of pronghorn skeletal parts at site 39FA83, and adjusted residuals and probability values for each. Based on data in Table 6.5*

Skeletal part	Left	Adjusted residual	p	Right	Adjusted residual	p
Mandible	18.8	0.26	0.397	18.2	0.28	0.390
Innominate	16.3	1.21	0.113	15.7	1.22	0.111
Scapula	24.4	0.12	0.452	23.6	0.12	0.452
P humerus	1.5	1.76	0.039	1.5	1.76	0.039
D humerus	28.5	0.71	0.239	27.5	0.73	0.233
P radius	27	0.29	0.386	26	0.30	0.382
D radius	22.9	0.03	0.492	22.1	0.03	0.488
P ulna	22.9	0.03	0.492	22.1	0.03	0.488
P metacarpal	19.3	2.61	0.004	18.7	2.62	0.004
P femur	8.7	1.13	0.129	8.3	1.16	0.123
D femur	8.1	1.07	0.142	7.9	1.09	0.138
P tibia	9.2	0.10	0.460	8.8	0.10	0.460
D tibia	25.5	1.95	0.026	24.5	1.98	0.024
P metatarsal	18.8	1.09	0.138	18.2	1.10	0.136

#### USING MNE VALUES TO MEASURE SKELETAL-PART FREQUENCIES

Other than Whites (1953b) suggestions regarding comparison of the frequencies or the spatial distributions of right-side skeletal parts and left-side skeletal parts, how have MNE data been used? As indicated earlier, they have been used largely by taphonomists who seek to discern if, and why, frequencies of skeletal parts diverge from a model of some number (usually the MNI of the collection under study) of complete skeletons. They have also been used to measure the degree of fragmentation evident in an assemblage. Procedures used to analyze skeletal-part frequencies are discussed first. Throughout, it is assumed that the principle of anatomical overlap or redundant skeletal parts has been used to measure MNE, and it is assumed that all skeletal specimens (e.g., diaphysis and epiphysis fragments of long bones) have been included.

Usually the number of complete skeletons to which observed skeletal-part frequencies are compared is that expected given the MNI for the taxon of concern. If the MNI is ten of an artiodactyl species, then there should be, for example, ten skulls, seventy cervical vertebrae, twenty humeri (= 10 left + 10 right), eighty first phalanges, and so on. Were one to graph the frequencies of the major categories of skeletal elements for

Table 6.7. *Frequencies of skeletal elements in a single generic artiodactyl skeleton*

Skeletal element	N	Skeletal element	N
Skull	1	Mandible	2
Cervical vertebra	7	Thoracic vertebra	13
Lumbar vertebra	6	Sacrum	1
Rib	26	Innominate	2
Scapula	2	Humerus	2
Radius	2	Ulna	2
Carpal	12	Metacarpal	2
Femur	2	Tibia	2
Tarsal	10	First phalanx	8
Second phalanx	8	Third phalanx	8

a single artiodactyl carcass (Table 6.7), a graph like the one shown in Figure 6.7 might be the result. (This type of graph is very similar to the one used by many early workers who had followed Whites [e.g., 1952] lead and interpreted skeletal-part frequencies.) Comparing the frequency of skeletal parts of a taxon in a prehistoric collection to that model would be difficult visually (using the graph) and also statistically (using a table of expected and observed frequencies). To simplify comparisons of observed frequencies with those manifested in the model, the model (expected frequencies) and the observed frequencies can be modified such that divergence of the latter from the former is made obvious. That is precisely what several analysts did beginning in the 1960s and 1970s.

#### *Modeling and Adjusting Skeletal-Part Frequencies*

Binford (1978, 1981, 1984; Binford and Bertram 1977) was not interested in the frequencies of left and right elements in a collection, or in the frequencies of third cervical and seventh thoracic vertebrae, or the like; such distinctions were unimportant to the questions he was asking of the remains of caribou (*Ranifer tarandus*) exploited by his Inuit informants nor were they relevant to the Paleolithic assemblages of faunal remains he was studying. Rather, he was interested in whether humeri preserved better than tibiae, whether Inuit hunters more often transported femora from kill sites to camp/consumption sites than they transported phalanges, and the like. Therefore, he divided MNE values for each anatomical part or portion by the number of times that part or portion occurs in one complete skeleton. Skulls were

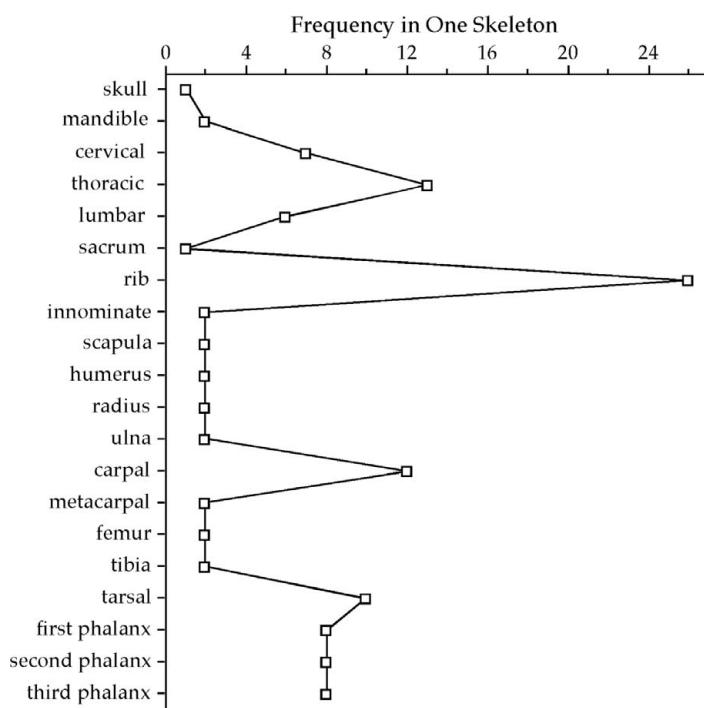


FIGURE 6.7. Frequencies of skeletal elements per category of skeletal element in a single artiodactyl carcass. Data from Table 6.7.

divided by one; mandibles and humeri and femora (etc.) by two; cervical vertebrae by seven; and so on (see Table 6.4 for various divisors). This standardized (or normed) the observed MNE counts to individual carcasses or skeletons. Each caribou skull represented one skeleton or individual, every two femora represented the equivalent of one skeleton whereas a single femur represented half of a skeleton or individual, every eight first phalanges of caribou represented one skeleton but every single first phalanx represented the equivalent of 1/8 or 0.125 individuals, and so on through the entire skeleton.

Binford (1984:50) ultimately referred to the skeletally standardized values as minimum animal units, or MAU values. Typically, MAU values themselves were normed by dividing all MAU values by the greatest observed MAU value in a particular collection and multiplying each resulting value by 100. Because the values produced ranged between 0 and 100 and were similar to percentages, they were (and are) sometimes referred to as %MAU values. Because the MNE values were all normed to the same scale, samples of faunal remains of quite different sizes could be compared graphically without fear of variation in sample size influencing the results. Thus, White

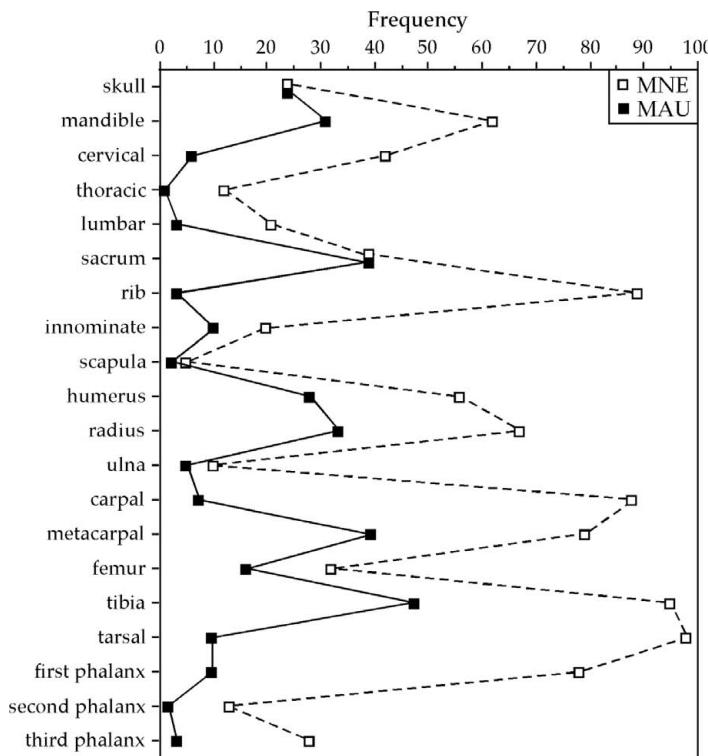


FIGURE 6.8. MNE and MAU frequencies for a fictional data set. Data from Table 6.8.

(1955, 1956) normed some of the assemblages he studied, as did others after him (e.g., Gilbert 1969, Kehoe and Kehoe 1960; Wood 1962, 1968), long before Binford (1978, 1981, 1984) popularized this analytical protocol.

Norming makes graphs of skeletal-part frequencies easier to interpret. Figure 6.8 presents the data from Table 6.8 in the same format as that in Figure 6.7, but both traditional MNE values per skeletal part or portion and MNE values standardized to a single artiodactyl skeleton, or MAU values, are presented. The data in Table 6.8 are fictional; a two-digit whole number was drawn from a table of random numbers and served as the MNE for a skeletal part or portion. Those values were divided by the values in Table 6.7 to generate the standardized, or MAU, values in Table 6.8. Notice the difference in the two sets of values plotted in Figure 6.8. Converting MNE values to MAU values mutes much of the variation between frequencies of skeletal parts and portions that is due to variation in how frequently a kind of part or portion is represented in a skeleton.

But the more important thing to realize is that a comparison of MNE values to a model skeleton is difficult to interpret, as exemplified in Figure 6.9, where the

Table 6.8. MNE and MAU frequencies of skeletal parts and portions. MNE frequencies were generated from a random numbers table

Skeletal element	MNE	MAU	Skeletal element	MNE	MAU
Skull	24	24	Mandible	62	31
Cervical vertebra	42	6	Thoracic vertebra	12	0.9
Lumbar vertebra	21	3.5	Sacrum	39	39
Rib	89	3.4	Innominate	20	10
Scapula	5	2.5	Humerus	56	28
Radius	67	33.5	Ulna	10	5
Carpal	88	7.3	Metacarpal	79	39.5
Femur	32	16	Tibia	95	47.5
Tarsal	98	9.8	First phalanx	78	9.75
Second phalanx	13	1.6	Third phalanx	28	3.5

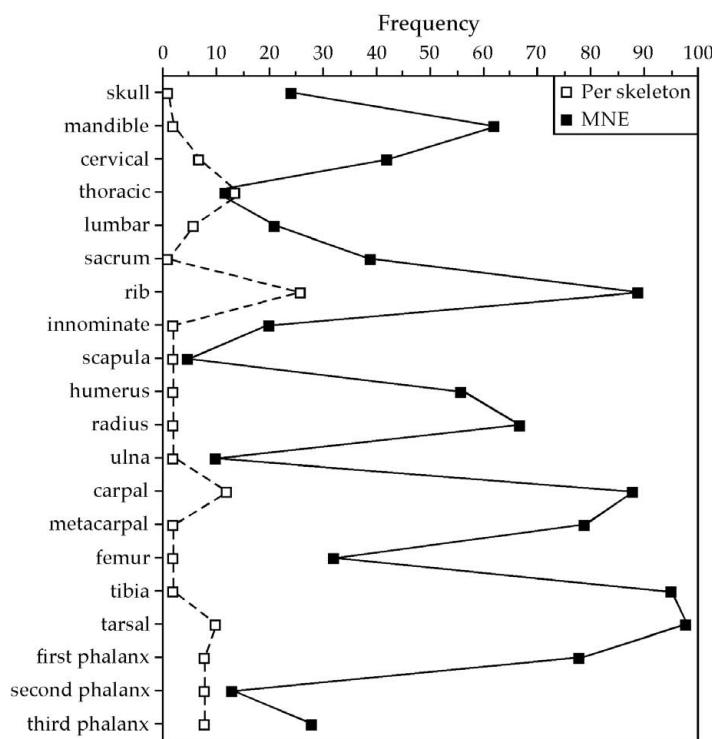


FIGURE 6.9. MNE values plotted against the MNE skeletal model. Data from Tables 6.7 and 6.8.

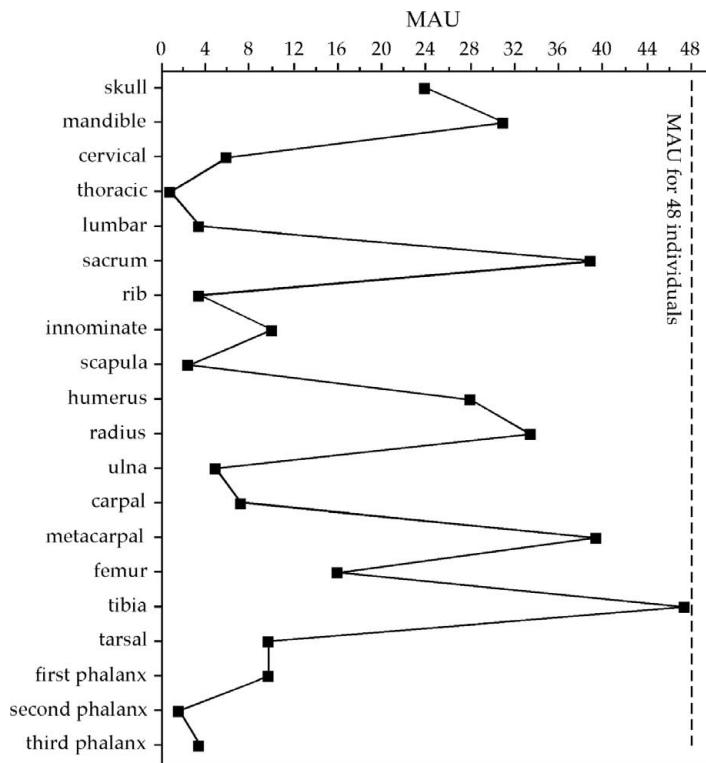


FIGURE 6.10. MAU values plotted against the MAU skeletal model (for forty-eight individuals). Data from Table 6.8.

MNE values from Table 6.8 are compared to the MNE model skeleton from Table 6.7 (graphed in Figure 6.7). MAU values present a more readily interpretable result, as exemplified by Figure 6.10, where MAU values from Table 6.8 are compared to the MAU model skeleton. Recall that MAU values are MNE values that have been standardized to a complete skeleton. That standardization process makes comparison of observed frequencies of skeletal parts and portions with a skeletal model more comprehensible because the MAU skeletal model sets the frequency of all skeletal parts and portions to the MNI observed in the total collection. If the MNI is seven, the MAU skeletal model is set to seven across all skeletal parts; if the MNI is forty-three, the MAU skeletal model is set to forty-three across all skeletal parts and portions; and so on.

A means to standardize MNE values related to the one described by Binford (1978, 1981, 1984; Binford and Bertram 1977) was designed with a different analytical goal in mind. Brain (1967, 1969, 1976) did not use the term MNE (nor did he use the term

MNI), though it is likely he used that quantitative unit to produce what he called the %survivorship of skeletal parts (Lyman 1994a). Given his descriptions of how he calculated %survivorship, Brain likely used the equation:

$$([MNE_i]100)/(MNI[\text{number of times } i \text{ occurs in one skeleton}]), \quad (6.1)$$

where  $i$  denotes a particular skeletal part or portion (e.g., proximal half of humeri, thoracic section of vertebral column). The denominator in this equation is the number of each skeletal portion to expect if 100 percent of them are in the collection, in light of the MNI for the collection. Thus, the denominator is equal to the maximum MNE in the assemblage. If there is an MNI of ten mammals, we would expect to find ten skulls, twenty humeri, and so on, depending on the taxon under consideration.

As it turns out, Brains %survivorship equation produces exactly the same value as Binfords %MAU. The latter is calculated with the equation (Binford 1978, 1981, 1984):

$$([MAU_i]100)/\text{maximum MAU in the assemblage}, \quad (6.2)$$

where  $i$  again denotes a particular skeletal part or portion. Note that  $MAU_i$  is determined with the equation:

$$(MNE_i)/\text{number of times } i \text{ occurs in one skeleton}. \quad (6.3)$$

Substituting Eq. 6.3 into Eq. 6.2,

$$([MNE_i/\text{number of times } i \text{ occurs in one skeleton}]100)/(\text{maximum MNE}_i/\text{number of times maximum}_i \text{ occurs in one skeleton}). \quad (6.4)$$

Given that the /number of times  $i$  occurs in one skeleton in the numerator and denominator cancel each other out, Eq. 6.4 is mathematically identical to Eq. 6.1.

Binford and Brain each determined a means to quantify skeletal parts and to scale them in such a manner as to allow graphing those values in forms that were easily interpreted. Various paleontologists derived similar equations that are in fact mathematically identical (Andrews 1990; Dodson and Wexlar 1979; Korth 1979; Kusmer 1990; Shipman and Walker 1980). It suffices here to describe one of them (see Lyman [1994b] for detailed discussion). Andrews (1990:45) suggested the equation:

$$R_i = N_i/(MNI)E_i, \quad (6.5)$$

where  $R_i$  is the relative (proportion) frequency of skeletal part  $i$ ,  $N_i$  is the observed frequency of skeletal part  $i$  in the assemblage, MNI is the minimum number of individuals in the assemblage, and  $E_i$  is the frequency of skeletal part  $i$  in one skeleton. (In practice, Andrews [1990] multiplies  $R_i$  by 100 to derive a percentage

Table 6.9. MAU and %MAU frequencies of bison (*Bison bison*) from two sites. Data for 32SL4 from Wood (1962); data for 24GL302 from Kehoe and Kehoe (1960)

Skeletal part	32SL4-MAU	32SL4-%MAU	24GL302-MAU	24GL302-%MAU
Skull	9	60	16	43
Mandible	11	73	37	100
Atlas	2	13	20	54
Axis	5	33	16	43
Cervical	2	13	15	41
Thoracic	1	7	12	32
Lumbar	1	7	6	16
Sacrum	0	0	8	22
Humerus	7	47	7	19
Radius	7	47	13	35
Ulna	11	73	10	27
Metacarpal	3	20	18	49
Innominate	1	7	12	32
Femur	2	13	15	41
Tibia	8	53	14	38
Astragalus	15	100	19	51
Calcaneum	8	53	18	49
Metatarsal	3	20	13	35
First phalanx	4	27	14	38
Second phalanx	9	60	18	49
Third phalanx	5	33	14	38

frequency.) Because  $N_i = MNE_i$ , and because  $(MNI)E_i =$  the expected number of parts were all of  $i$  present given MNI, then Eq. 6.5 is mathematically the same as Eqs. 6.1 and 6.4.

Recall that norming MAU values to %MAU, or calculating %survivorship, allows one to graphically compare samples of different sizes. Deciphering the significance of a comparison of MAU values determined for one collection that contains remains of five individuals with another collection that contains remains of twenty individuals would be difficult without norming both to the same scale. Table 6.9 presents data from two sites, one with remains of an MNI of fifteen bison (*Bison bison*) (Wood 1962) and the other with remains of an MNI of thirty-seven bison (Kehoe and Kehoe 1960). If the unnormed MAU values are graphed together, it is difficult to discern what is happening (Figure 6.11). But if both sets of MAU values are normed to %MAU values and graphed, then it is much easier to discern similarities and differences in the

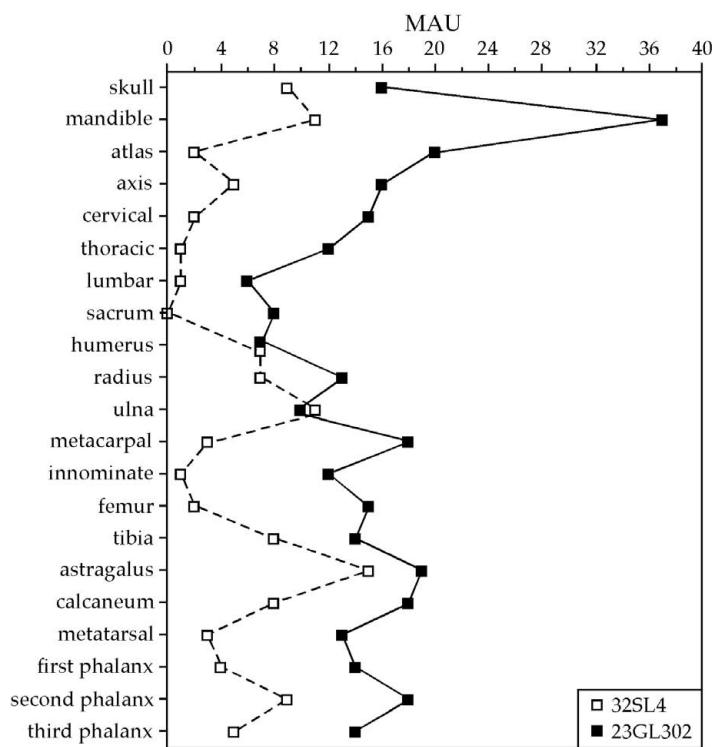


FIGURE 6.11. MAU values for two collections with different MNI values. MNI at 32SL4 is 15; MNI at 23GL302 is 37. Compare with Figure 6.12. Data from Table 6.9.

frequencies of skeletal parts and portions (Figure 6.12). Such norming, however, gives values that some statisticians suggest cannot be analyzed statistically; the influence of sample size is, for example, masked by such norming and will influence statistical results as a consequence.

MNE originally (if implicitly) formed the basis of the MNI quantitative unit. As research interests shifted from taxonomic abundances to include consideration of skeletal-part abundances, MNE became a unit in need of explicit recognition. And that is in fact what it received beginning in the 1960s and especially the 1970s. With explicit recognition came consideration of how to operationalize MNE. It would be interesting to review the numerous discussions of how to determine MNI that appeared in the 1950s through 1980s with the goal of ascertaining if the commentators worried as much about operationalizing MNI as those who in the 1980s and 1990s have worried about operationalizing MNE.

MNE became a very popular and much used quantitative unit after about 1980. It was difficult to read an article on paleozoology (especially in the zooarchaeology

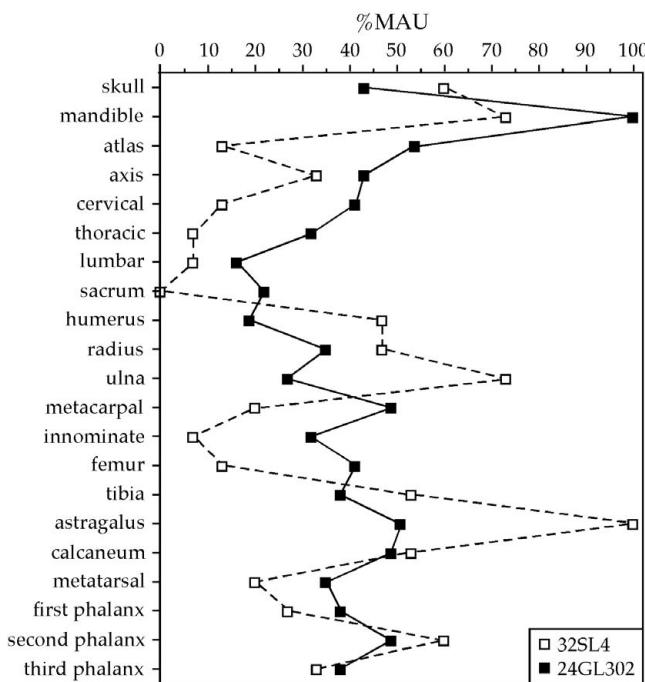


FIGURE 6.12. %MAU values for two collections with different MNI values. MNI at 32SL4 is 15; MNI at 24GL302 is 37. Compare with Figure 6.11. Data from Table 6.9.

literature) without encountering it or one or more of its derivatives (MAU, %MAU, %survivorship, etc.), likely because of the increased frequency of detailed taphonomic analyses that centered around questions thought to be answerable by analyses of skeletal-part frequencies. Interestingly, the quantitative units derived from MNE may have a utility that could serve a long sought after analytical goal. That goal is to measure the average or overall completeness of the skeletons represented in a collection. Are those skeletons all more or less complete, or are they relatively incomplete? It is to measuring that variable that we next turn.

#### MEASURING SKELETAL COMPLETENESS

In the 1950s, Shotwell (1955, 1958) developed a method that he thought would allow a paleozoologist to separate the remains of animals originating in one biological community from the remains of animals that originated in another community. He referred to the local community the one in which the collection locality was located as the *proximal community*, and any other community that might have contributed

taxa (but nonlocal) as the *distal community*. Shotwell reasoned that members of the proximal community would be more skeletally complete than members of the distal community.

Commentators identified taphonomic difficulties with sorting out the members of the two communities (e.g., Clark and Guensburg 1970; Dodson 1973; Voorhies 1969; Wolff 1973). These included the assumption that skeletons of individuals originating near the site of accumulation and deposition would be more complete than those of individuals that originated some distance away. This assumption presumed that bone accumulation processes (mechanisms, such as fluvial transport, and agents, such as carnivores) would operate according to a principle of distance decay—the greater the distance away, the fewer of an organism's remains that would be transported to and deposited at the (future) site of recovery. Shotwell (1958:272) stated that the community with the greater relative [skeletal] completeness is the one nearest to the site of deposition and is therefore referred to as the proximal community. We now know that the mechanisms and agents that accumulate faunal remains display no consistent or universal distance decay pattern. Sometimes they do, sometimes they do not. Shotwell's suggestion is best considered as a hypothesis that warrants examination on a case-by-case basis.

Shotwell's method involves determination of MNI, and then calculating the corrected number of specimens per individual (CSI). The CSI is the index used to determine whether a taxon represents a member of the proximal or distal community. Although Shotwell (1955, 1958) used the terms specimens and elements interchangeably in his discussion, what he had in mind was MNE values rather than NISP values. He wrote his formula as:

$$\text{CSI} = 100(\text{NISP})/\text{number of elements per skeleton}, \quad (6.6)$$

where CSI is the corrected number of specimens (per individual), and the number of elements per skeleton is the number that could be identified in a complete skeleton, excluding ribs and vertebrae. Because the *average skeletal completeness per individual* per taxon is the desired measure, the denominator should be multiplied by the MNI of the taxon under study. Because Shotwell was dealing with skeletal elements that were anatomically complete, his formula for measuring skeletal completeness can be rewritten more completely as:

$$\begin{aligned} \text{CSI}_i &= 100(\text{MNE})/\text{MNE per complete skeleton} \\ &\quad (\text{minus vertebrae and ribs}) \times [\text{MNI}]. \end{aligned} \quad (6.7)$$

For one standard artiodactyl skeleton as in Table 6.7, the denominator would be sixty-five. This step accounts for the fact that a standard artiodactyl, for instance,

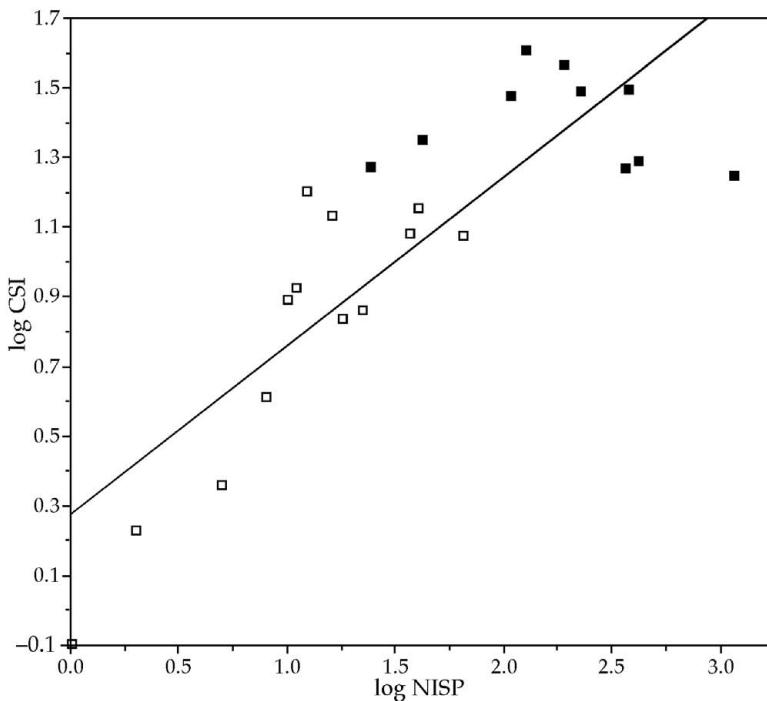


FIGURE 6.13. Relationship between Shotwell's CSI (or skeletally normed NISP/MNI ratio) per taxon and NISP per taxon for the Hemphill paleontological mammal assemblage. Taxa assigned to the proximal community by Shotwell are represented by filled squares; they have high NISP/MNI ratios, but also greater sample size than taxa assigned to the distal community (unfilled squares). Simple best-fit regression line ( $Y = 0.276X^{0.48}$ ) shown for reference ( $r = 0.85$ ,  $p < 0.001$ ). Data from Shotwell (1958).

has a different number of identifiable elements than a standard perissodactyl (fewer phalanges than an artiodactyl), or a standard canid (more metapodials and phalanges than an artiodactyl) (Table 6.4). It is important to note that in both the numerator and the denominator, MNE is the total number of elements present in a collection *regardless of which elements are represented*.

Grayson (1978b) argued that Shotwell's method was flawed for statistical reasons, regardless of the history of accumulation of faunal remains. Grayson showed that the calculation of skeletal completeness using Shotwell's method produces a measure of sample size. Using Shotwell's original formula (Eq. 6.6), CSI is a skeletally normed ratio of NISP/MNI. As Grayson (1978b) showed, the ratio NISP/MNI varies with NISP. Figure 6.13 shows CSI and NISP values for one of Shotwell's (1958) assemblages plotted against one another. As NISP (sample size) increases, so too does the value of the ratio of NISP/MNI. There is an autocorrelation between

the two variables because NISP occurs on both sides; taxa with larger samples will appear to be more skeletally complete because of the relationship shown in Figure 2.4.

Thomas (1971) adapted Shotwell's method to an archaeological setting. Thomas (1971:367) reasoned that the assumption of an archaeologist was not that the taxa from a proximal community would be more skeletally complete than those from a distal community, but rather that skeletons of taxa exploited by humans would be more skeletally incomplete than skeletons of animals that died naturally on the site; "The primary assumption for the [zoo]archaeologist to evaluate is that the dietary practices of man tend to destroy and disperse the bones of his prey species." This may seem to be a reasonable assumption, but it is taphonomically naïve for the same reasons that Shotwell's method is. Thomas's (1971) alteration of Shotwell's method is flawed for the same reason that Shotwell's original method is flawed. Thomas did not really modify Shotwell's method, but instead used exactly the same reasoning and formula to measure skeletal completeness; what he did different than Shotwell was to argue that the least skeletally complete taxa (Shotwell's distal community) were the ones that humans had accumulated and deposited. Figure 6.14 is a graph of the data from one site analyzed by Thomas (1971) in the same form as Shotwell's data is graphed in Figure 6.13. Again, the relationship between CSI (or the standardized ratio of NISP/MNI) is strongly correlated with NISP. And, it is obvious that the least skeletally complete taxa not only are the ones that Thomas suggests were accumulated and deposited by human predators, but those same taxa have the smallest sample sizes.

### *A Suggestion*

Perhaps because of the serious statistical weaknesses of Shotwell's method, whether used as he originally intended or as Thomas (1971) modified it, no paleozoologist has used it since the 1970s. Reitz and Wing's (1999:255) nonjudgmental mention of the method is the only reference to it published since 1980 of which I am aware. Shotwell's method failed for taphonomic reasons and also for statistical reasons. What no one has previously noted is that the method did not directly take account of variation in the frequencies of individual skeletal parts and portions; rather, it merely calculated an average skeletal completeness based on whatever skeletal parts and portions per individual were represented. The skeletal completeness index was calculated the same way regardless of the skeletal parts and portions present in the

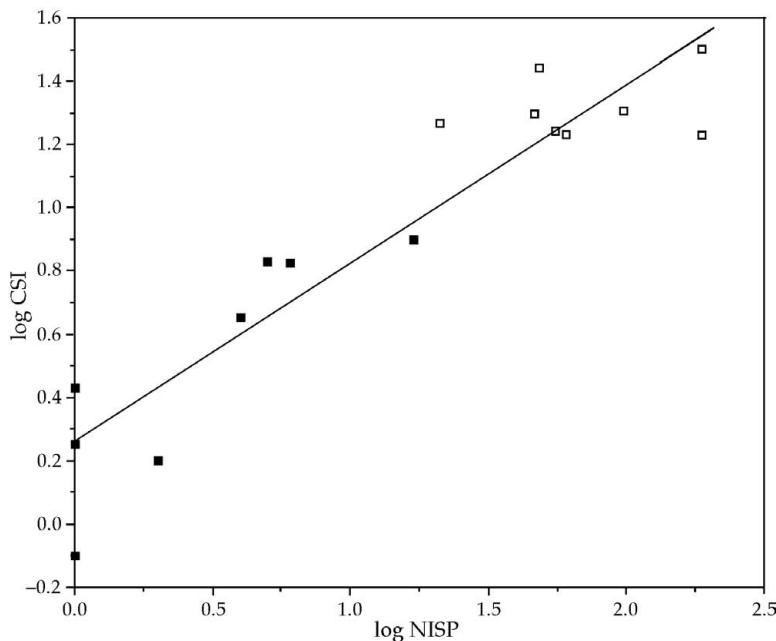


FIGURE 6.14. Relationship between Thomas's CSI per taxon and NISP per taxon for the Smoky Creek zooarchaeological mammal collection. Taxa thought to be accumulated and deposited by humans are represented by filled squares; they have low NISP/MNI ratios, but also smaller sample size than taxa thought to have been naturally accumulated and deposited (unfilled squares). Simple best-fit regression line ( $Y = 0.262X^{0.564}$ ) shown for reference ( $r = 0.93$ ,  $p < 0.001$ ). Data from Thomas (1971).

collection. (This is analogous to the flaw with skeletal mass allometry described in Chapter 3.)

The manner in which MNE has been used to examine skeletal completeness suggests a solution to some of the problems with Shotwell's original method. Table 6.10 provides MNE values and MAU values for twenty skeletal parts and portions of two taxa, each of which has an MNI of 10. Taxon 1 has relatively even skeletal-part frequencies (after Faith and Gordon 2007) measured as MAU values (Shannon's  $e = 0.999$ ), whereas taxon 2 has less even skeletal-part frequencies ( $e = 0.962$ ). Individuals of taxon 1 are more skeletally complete, on average, than individuals of taxon 2 (Figure 6.15). One might argue, then, that taxon 1 comprises the proximal community whereas taxon 2 comprises the distal community. If there are  $> 2$  taxa, follow Shotwell's lead (perhaps) and assign all taxa with  $e >$  average (or some other value) to the proximal community and taxa with  $e <$  average to the distal community. This version as well as Shotwell's original method assumes taxa

Table 6.10. *Skeletal-part frequencies (MNE and MAU) for two taxa of artiodactyl. MNI = 10 for both. Data are fictional*

Skeletal element	Taxon 1–MNE	1–MAU	Taxon 2–MNE	2–MAU
Skull	10	10	5	5
Mandible	19	9.5	8	4
Cervical vertebra	68	9.7	42	6
Thoracic vertebra	120	9.2	30	2.3
Lumbar vertebra	56	9.3	31	5.2
Sacrum	8	8	3	3
Rib	245	9.4	200	7.7
Innominate	18	9	2	1
Scapula	19	9.5	5	2.5
Humerus	20	10	18	9
Radius	18	9	16	8
Ulna	17	8.5	14	7
Carpal	115	9.6	42	3.5
Metacarpal	16	8	8	4
Femur	19	9.5	20	10
Tibia	19	9.5	15	7.5
Tarsal	92	9.2	65	6.5
First phalanx	78	9.8	44	5.5
Second phalanx	75	9.4	32	4
Third phalanx	72	9	16	2

will *not* occur in both the proximal and the distal communities, which may be false.

In the preceding paragraph MAU values were used rather than MNE values to measure the evenness of skeletal-part frequencies because the former provide the model baseline. Any number of anatomically complete skeletons would produce a Shannon's  $e = 1.0$  using MAU values; any number of anatomically complete skeletons would produce a Shannon's  $e = 0.862$  using MNE values (the MNE values in Table 6.7 were used to generate this  $e$  value), making any observed value more difficult to interpret (an observed  $e$  would range from 0 to 1.0 rather than from 0 to 0.862). The example in Table 6.10 and Figure 6.15 is fictional. With real data, a critical early step is to determine if the numbers of left and the numbers of right elements of bilaterally paired bones are not significantly different (e.g., Table 6.6). If they are significantly different, then measures of skeletal-part abundances calculated as MAU

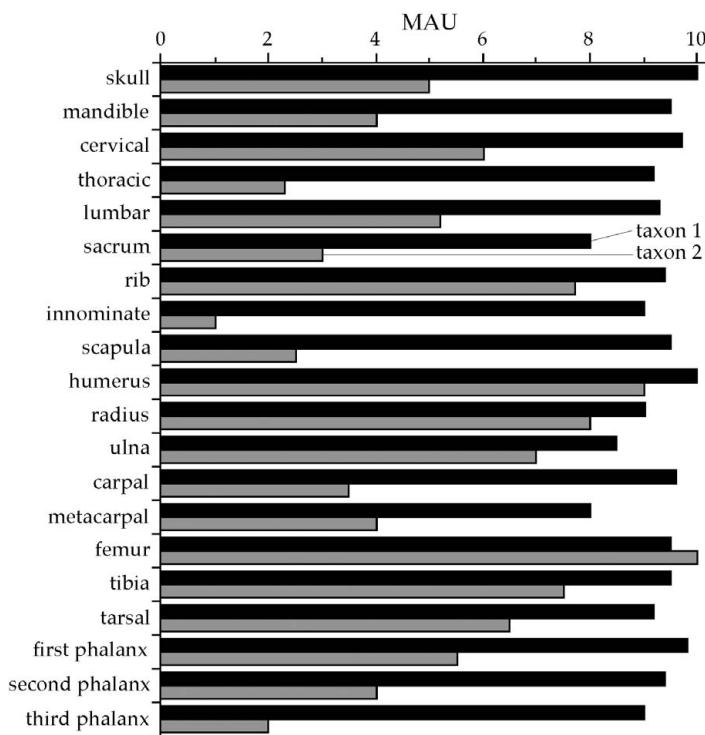


FIGURE 6.15. Bar graph of frequencies of skeletal parts (MAU) for two taxa. Data from Table 6.10.

will be meaningless with respect to the completeness of individual skeletons. Recall that MAU values ignore left and right distinctions and divide the observed MNE by the number of times a skeletal part or portion occurs in one skeleton. If there are major discrepancies in the frequencies of left and right elements, then the MNI (maximum number of, say, left or right elements = number of skeletons) will be considerably larger than any MAU value (say,  $[lefts + rights]/2$ ).

The procedure of measuring skeletal completeness using the evenness of skeletal parts measured as MAU values described in the preceding paragraph should not be adopted uncritically (if at all). The evenness of skeletal parts may be a function of sample size (Chapter 5). And recall the other problems that attend MNE – influences of aggregation and definition. These problems also influence MAU and %MAU values. Graphs such as that in Figure 6.15 may help evaluate the degree of skeletal completeness. But if NISP and MNE (or MAU) per skeletal part are correlated, then the two quantitative units provide redundant information on skeletal-part

frequencies. As we saw in Chapter 2, NISP is not afflicted by problems of aggregation or definition, so evaluating skeletal completeness can be done with NISP to avoid the problems with MNE and MAU (see Grayson and Frey [2004] for additional discussion).

Whether taxa with complete skeletons derived from a local (proximal) community, or were naturally deposited, and whether taxa with incomplete skeletons derived from a distant (distal) community, or were deposited by particular bone accumulating processes, are different questions. Deciding what that completeness (or lack thereof) means in terms of taphonomy requires more actualistic research. If Shotwell and Thomas were correct, and even if they were not, the concept of skeletal-part frequencies suggests a technique to measure skeletal completeness in a much more anatomically realistic way than that proposed by Shotwell. The new technique comes from explicit recognition of MNE as a quantitative unit, second, that MNE is derived (inherently problematic), and third, that NISP and MNE are often correlated. The solution is to alter the value plotted on the horizontal axis of Figure 6.15 from MAU to NISP.

Recall that the NISP and MNE value pairs for deer remains from the Meier site are strongly correlated, as are those values for wapiti remains. The frequency distributions of each suggest that the NISP values provide ordinal scale data on skeletal-part frequencies (Figures 6.3 and 6.4). Are deer or are wapiti more completely represented skeletally? The Shannon evenness index for the NISP per skeletal part of the two taxa are: deer,  $e = 0.993$  (heterogeneity [ $H$ ] = 3.114); wapiti,  $e = 0.925$  ( $H = 2.901$ ). Deer skeletons are a bit more completely represented than are wapiti skeletons. This is perhaps explicable by the fact that deer are smaller than wapiti and some carcasses of adult deer might be transported as a single unit (they must be divided into smaller parts [usually anatomical quarters] in many instances), wapiti are considerably larger than deer and must always be divided into smaller portions for transport purposes. Butchering (reduction) of carcasses for transport is likely to be more extensive and result in more culling and discard of wapiti bones at a kill site than is butchering of deer.

There is yet another possible way to compare frequencies of skeletal parts of two assemblages (whether two different taxa in the same collection or the same taxon in two different collections). The technique is to calculate a  $\chi^2$  statistic, and if it is significant and thus indicates a significant divergence from random frequencies, then calculate adjusted residuals for each value to determine which frequencies diverge from randomness. Comparison of the NISP per skeletal-part data for deer and wapiti in Table 6.3 indicates the two sets of values are significantly different ( $\chi^2 = 90.41$ ,

Table 6.11. *Expected (EXP) frequencies of deer and wapiti remains at Meier, adjusted residuals (AR), and probability values for each (p). Based on data in Table 6.3*

Skeletal part	Deer EXP	Wapiti EXP	Deer AR	Wapiti AR	Deer p	Wapiti p
mandible	176.3	43.7	2.741	-2.746	<0.01	<0.01
atlas	37.7	9.3	2.316	-2.324	<0.05	<0.05
axis	19.2	4.8	-0.103	0.102	>0.1	>0.1
cervical	77.7	19.3	-0.180	0.180	>0.1	>0.1
thoracic	81.7	20.3	-1.597	1.599	>0.1	>0.1
lumbar	111.4	27.6	-1.604	1.608	>0.1	>0.1
rib	226.0	56.0	-0.778	0.779	>0.1	>0.1
innominate	131.4	32.6	-0.280	0.281	>0.1	>0.1
scapula	68.1	16.9	1.350	-1.348	>0.1	>0.1
humerus	141.0	35.0	1.743	-1.744	>0.05	>0.05
radius	163.5	40.5	0.090	-0.090	>0.1	>0.1
ulna	95.4	23.6	1.543	-1.546	>0.1	>0.1
metacarpal	133.8	33.2	-0.159	0.159	>0.1	>0.1
femur	95.2	24.8	-2.153	2.101	<0.05	<0.05
patella	12.0	3.0	0.649	-0.647	>0.1	>0.1
tibia	176.3	43.7	2.394	-2.393	<0.05	<0.05
astragalus	116.2	28.8	2.293	-2.297	<0.05	<0.05
calcaneum	141.9	35.1	3.301	-3.313	<0.05	<0.05
naviculo-cuboid	76.1	18.9	2.585	-2.579	<0.05	<0.05
metatarsal	153.1	37.9	-1.880	1.888	>0.05	>0.05
first phalanx	248.5	61.5	-3.656	3.662	<0.01	<0.01
second phalanx	181.1	44.9	-3.708	3.985	<0.01	<0.01
third phalanx	80.1	19.9	-1.298	1.296	>0.1	>0.1

$p < 0.001$ ). Calculation of expected values and adjusted residuals indicate that nine of the twenty-three included skeletal parts differ significantly between the two taxa in terms of their abundances (Table 6.11). Relative to wapiti remains, deer mandibles, atlas vertebrae, tibiae, astragali, calcanei, and naviculo-cuboids are overrepresented whereas deer femora, first phalanges, and second phalanges are under represented. Why this is the case is a taphonomic question, but quantitative analysis has identified the significant variation and indicates those aspects of the data requiring taphonomic analysis. One of the analytical avenues that might be explored in trying to determine why variation in NISP occurs between the two taxa is variation in fragmentation, which brings us to methods for measuring fragmentation.

## MEASURING FRAGMENTATION

Each distinct kind of skeletal element can be conceived as a model of how a particular kind of “natural” biological thing looks. If a specimen of a skeletal element – femur, atlas vertebra, or third upper molar – is anatomically incomplete, it is not biologically natural but instead is fragmentary (relative to the model). Even if the entire element is represented, it may be in unnatural pieces or fragments (just as the bones and teeth of a complete skeleton might be disarticulated and dispersed in a unnatural arrangement). Thus, individual, anatomically complete skeletal elements can be conceived of as not only whole or complete, but as single or individual discrete entities (ignoring for sake of discussion whether a tooth embedded in a mandible is a separate, distinct, discrete skeletal element or not). Given the model of natural whole discrete skeletal elements, an obvious quantitative measure is the number of pieces (fragments) that each skeletal element has been broken into or is represented by.

Paleozoologists have worried about the degree of fragmentation of faunal remains for decades, as evidenced by their worries about intertaxonomic variation in fragmentation differentially skewing NISP measures of taxonomic abundances (Chapter 2). Tallying MNE frequencies per taxon escapes that problem, but introduces the problems attending derivation of MNE (aggregation, sample size, definition). Taphonomic questions about fracturing agents and processes have resulted in some innovations in measuring fragmentation. Simply because taxon A has a greater NISP value than taxon B does not mean that the remains of taxon A are more fragmentary than those of taxon B. How might we determine which taxon’s remains are the more anatomically complete and less fractured, and which taxon’s are more anatomically incomplete and more fragmentary?

### *Fragmentation Intensity and Extent*

Klein and Cruz-Uribe (1984) use the ratio NISP/MNI per skeletal element to measure fragmentation for each taxon, but there is a potential problem with this measure. NISP is the number of identified specimens, and a specimen is a bone or tooth or *fragment thereof*. The last two words are emphasized for one simple reason. If, say, many of the skeletal elements of taxon A are anatomically complete but a few of each were broken into many (small) fragments, then the ratio NISP/MNI for that taxon may be the same as that for taxon B all the remains of which are (large) fragments. This suggests that there are two dimensions of fragmentation. The *extent*

*of fragmentation* is the dimension that signifies the proportion or percentage of specimens in a collection that are anatomically incomplete, or its complement, the percent of NISP that comprise anatomically complete specimens, or %whole (Lyman 1994b, 1994c). The *intensity of fragmentation* signifies how small fragments are or how many pieces a kind of skeletal element has been broken into on average (Lyman 1994b, 1994c).

To calculate %whole or %fragmentary, tally up NISP for a taxon. Then, tally the number that are anatomically complete or whole (how many are actually skeletal elements rather than fragments of elements); this number will likely be (sometimes quite significantly) smaller than the number of fragmentary specimens. Divide the number of whole specimens by the total number of specimens (and multiply by 100) to derive the %whole; or, subtract the number of whole specimens from the total number of specimens, and divide the resulting number (number of fragments) by the total NISP (and multiply by 100) to derive the %fragmentary.

In the collection of faunal remains from the sample of owl pellets (Table 2.9), skulls of *Microtus* are not always complete. The total NISP of *Microtus* skulls is 110; the number of complete skulls is 103; the %whole of *Microtus* skulls is 93.6 percent. In that same collection, the total NISP of *Peromyscus* skulls is 206; the number of complete skulls is 115; the %whole of *Peromyscus* skulls is 55.8 percent. The extent of fragmentation of *Microtus* skulls is considerably less than is the extent of fragmentation of *Peromyscus* skulls. Why this difference should exist given the identical taphonomic histories of the two may now be explored; it likely is a result of *Peromyscus* skulls being much more fragile and of much more gracile structure than *Microtus* skulls, which are larger and more robust.

#### *The NISP:MNE Ratio*

But what if, in a case like the skulls of the two taxa of rodents just described, the specimens of the taxon with greater %whole are smaller than the fragments of the taxon with lower %whole? This concerns fragmentation intensity and it is measured as the ratio of anatomically incomplete specimens to the MNE represented by those specimens. (To calculate this ratio one must assume MNE values are not influenced by aggregation, sample size, or definition. Alternatively, one could assume the influences on MNE are randomly distributed across the collections compared such that they do not skew the values in such a way as to influence statistical interpretation.) Anatomically complete specimens are not included in the calculation because (i) when they are included they decrease the ratio because they increase both values

Table 6.12. Ratios of NISP:MNE for four long bones of deer in two sites on the coast of Oregon State. Data from Lyman (1995b)

Skeletal element	Umpqua/Eden site	Seal Rock site
Humerus	1.14	1.75
Radius	2.00	1.89
Femur	1.50	2.67
Tibia	2.10	2.33

of the NISP:MNE ratio equally, and (ii) the intensity of fragmentation is meant to capture the variable of fragment size. With respect to the first point, in an assemblage of anatomically incomplete specimens, if NISP = 10 and MNE = 5, then the ratio is 2:1. If two anatomically complete specimens are included, NISP = 12, MNE = 7, and the ratio is reduced to 1.71:1. The more anatomically complete specimens, the less the difference between NISP and MNE. With respect to the second point – NISP:MNE measures fragment size – higher ratios suggest smaller fragments. A ratio of 2:1 suggests elements were basically broken in half; a ratio of 15:1 suggests elements were almost pulverized.

Let's say we want to determine if the fragmentation of bones of a taxon differs across two assemblages. (The remains of two taxa can be compared using the same technique.) I did this for the four major long bones of deer (*Odocoileus* sp.) using the remains from two sites on the coast of the state of Oregon (Lyman 1991, 1995b). The ratios (Table 6.12) suggest that, overall, long bones were less intensively fractured (broken into larger pieces) at the Seal Rock site than they were at the Umpqua/Eden site. The average ratio at Seal Rock is 1.68 compared to an average ratio at Umpqua/Eden of 2.16. Determination of the reason for this apparent difference in fragmentation intensity requires other sorts of analyses. The NISP:MNE ratio allows one to rank-order the elements from most intensively broken to least intensively broken. For Seal Rock, that order is femur, tibia, radius, humerus; for Umpqua/Eden, that order is tibia, radius, femur, humerus. Of course, variation in the NISP:MNE ratio can be assessed and compared across different taxa as well.

Ratios of NISP:MNE have been used by zooarchaeologists in both the New World (e.g., Wolverton 2002) and the Old World (e.g., Munro and Bar-Oz 2005) to measure the intensity of fragmentation. But a property of the ratio needs to be identified. The ISP of NISP concerns identified specimens, and to be identified a specimen must retain sufficient anatomical and taxon-specific features to be identified. As

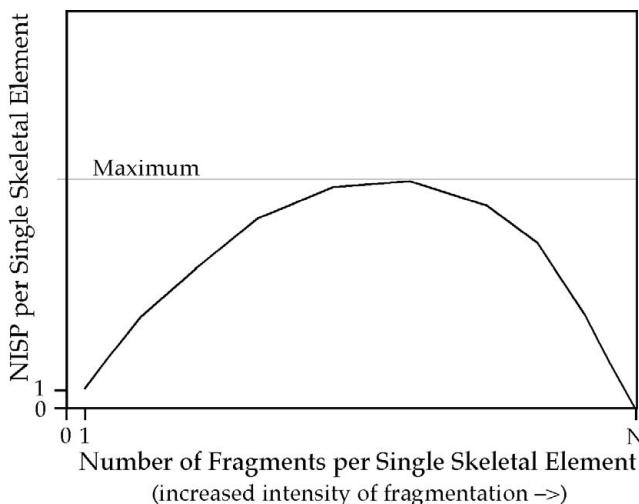


FIGURE 6.16. Model of the relationship between fragmentation intensity and NISP. The value of Maximum and of  $N$  are unknown. Modified from Marshall and Pilgram (1993).

elements are broken into successively smaller and smaller fragments, the resulting pieces become successively less likely to retain sufficient landmarks to permit their identification. Thus, as Marshall and Pilgram (1993) pointed out some years ago, fragmentation has the effect of first increasing the NISP represented by pieces of any given skeletal element, but as fragmentation intensity increases beyond some as yet unknown level of intensity, NISP will level off and then decrease because the fragments are becoming so small as to be unidentifiable (Figure 6.16). This may be an interpretively treacherous property of fragmentation if some kinds of skeletal elements are represented by only one identifiable fragment and numerous unidentifiable small fragments. Each identifiable fragment will represent an MNE of 1, and so the ratio for these would be 1:1, or 1, but slightly larger and (thus) identifiable fragments will cause the ratio to be  $> 1.0$ .

The preceding leads to another observation. There is some threshold of fragment size controlling whether or not a fragment is identifiable. If the fragment is smaller than the threshold size, it is not identifiable. For the sake of illustration, grant the (no doubt somewhat unrealistic) assumption that for any given skeletal element, the element can be broken into some number of pieces of equal size. If we set that threshold number of pieces at fifteen, such that if, say, a humerus is broken into fifteen pieces, all can be identified (to skeletal element and to taxon), then if that humerus is broken into sixteen pieces, none of them will be identifiable (to skeletal element or to taxon). The implication of this observation is graphed in Figure 6.17. That

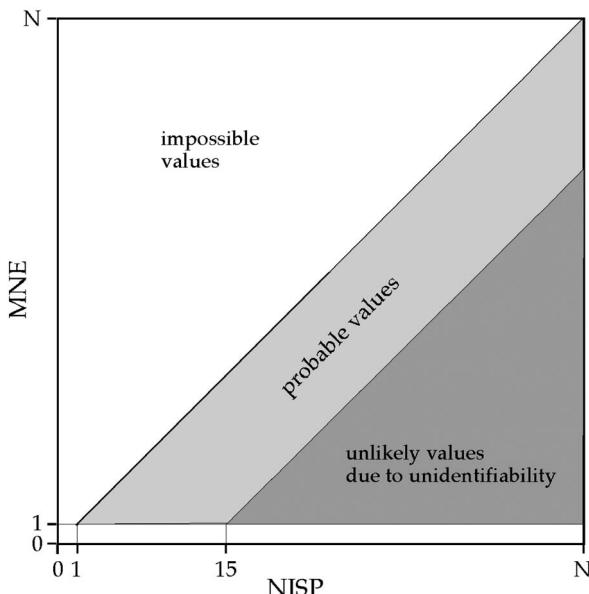


FIGURE 6.17. Model of the relationship between NISP and MNE. Observed values will always fall on or below the diagonal ( $\text{NISP} \geq \text{MNE}$ ), but not infinitely far below the diagonal because fragments that are too small will not be identifiable. After Lyman (1994b).

graph suggests there will be a relatively strong correlation between MNE and NISP per skeletal element simply because of constraints on the total NISP (identifiable fragments) that can be generated from any given set of fragments of multiple skeletal elements. Again, that these two variables are often strongly correlated should come as no great surprise given that whatever the greatest (left or right) MNE was for a taxon, that value was the MNI for that taxon. Nevertheless, we need to examine the relationship of these two variables in depth.

## DISCUSSION

There have been subsequent phases to the history of MNE. One later phase was the derivation of MAU and the related %MAU, along with the mathematically equivalent %survivorship. Those units served as the basis for a large volume of research on skeletal-part frequencies, and because they are ultimately based on MNE, they also prompted a plethora of research projects on how to determine MNE in the most accurate way possible. The former focused on what frequencies of skeletal

Table 6.13. *African bovid size classes. After Brain (1974, 1981) and Klein\* (1978, 1989)*

Bovid size class	Live weight
I (small)	0–23 kilograms (0–50 pounds)
II (small medium)	23–84 kilograms (50–200 pounds)
III (large medium)	84–296 kilograms (200–650 pounds)
IV (large)	>296 kilograms (>650 pounds)
V (very large)*	?

parts – normed to the model of the MNI of complete skeletons – meant with respect to taphonomic agents and processes of dispersal (e.g., fluvial winnowing), accumulation (e.g., nutritional value to carnivores and hominids), and destruction (e.g., carnivore gnawing) (see Lyman [1994c] for discussion of methods, and see the *Journal of Taphonomy* [2004: vol. 2] for more recent considerations). Earlier in this chapter efforts to derive the most accurate MNE values possible were outlined. Subsequent to those efforts, a new chapter or phase to the history of MNE was written.

Grayson and Frey (2004) recently showed that the relationship between NISP per skeletal element and MNE per skeletal element is strong; the two variables are often tightly correlated. As indicated earlier in this chapter, that a strong relationship exists between these two variables shouldn't really be a surprise, given that NISP and MNI are often tightly correlated and that MNI is operationalized as the largest value of MNE (of left or right specimens) per taxon, and given the model in Figure 6.17. But the fact that people grappled with MNE and its derivatives for more than twenty-five years before the statistical and analytical significance of the correlation of MNE and NISP was identified suggests that this significance should be illustrated. The relationship is shown in Figures 6.1 and 6.2, but it warrants additional discussion given the analytical weight placed on MNE abundances over the past two decades.

Researchers have published both NISP and MNE data for collections from diverse geographic locations and temporal periods. In one such presentation, Marean and Kim (1998) described frequencies of skeletal parts for an assemblage of remains representing several species of medium-small (size class II [Brain 1981]) bovids and cervids. (Bovid size classes used in much of the following are summarized in Table 6.13.) The remains originate from a Mousterian (Middle Paleolithic) deposit in Kobeh Cave, located in the Zagros Mountains of Iran. The data are summarized

Table 6.14. NISP and MNE frequencies of skeletal parts of bovid/cervid size class II remains from Kobeh Cave, Iran.  
Data from Marean and Kim (1998)

Skeletal part/portion	NISP	MNE
Horn	43	20.00
Skull	60	19.00
Mandible	75	22.00
Upper teeth	46	29.60
Lower teeth	82	21.86
Atlas	5	0.90
Axis	1	0.40
Cervical	24	8.35
Thoracic	28	11.30
Lumbar	28	8.60
Sacrum	2	1.90
Ribs	266	30.80
Humerus	404	63.80
Radius	336	47.25
Ulna	127	25.10
Carpal	14	11.50
Metacarpal	319	37.95
Innominate	53	11.30
Femur	478	62.90
Tibia	665	95.70
Astragalus	3	3.00
Calcaneum	13	4.90
Metatarsal	307	35.85
Tarsal	10	7.85
Phalange	102	24.90
Sesamoid	7	6.00

in Table 6.14, and graphed in Figure 6.18. The two variables are strongly correlated ( $r = 0.94$ ,  $p < 0.0001$ ), suggesting that the information regarding skeletal-part frequencies provided by MNE is virtually identical to that provided by NISP.

The close relationship between NISP and MNE is widespread. Enloe et al. (2000) described a large sample of saiga antelope (*Saiga tatarica*) remains from Prolom II Cave, located on the Crimean Peninsula in the Ukraine. The material dates to the Mousterian cultural period (Table 6.15). NISP and MNE are strongly correlated ( $r = 0.92$ ,  $p < 0.0001$ ); the MNE values are redundant with the NISP values with

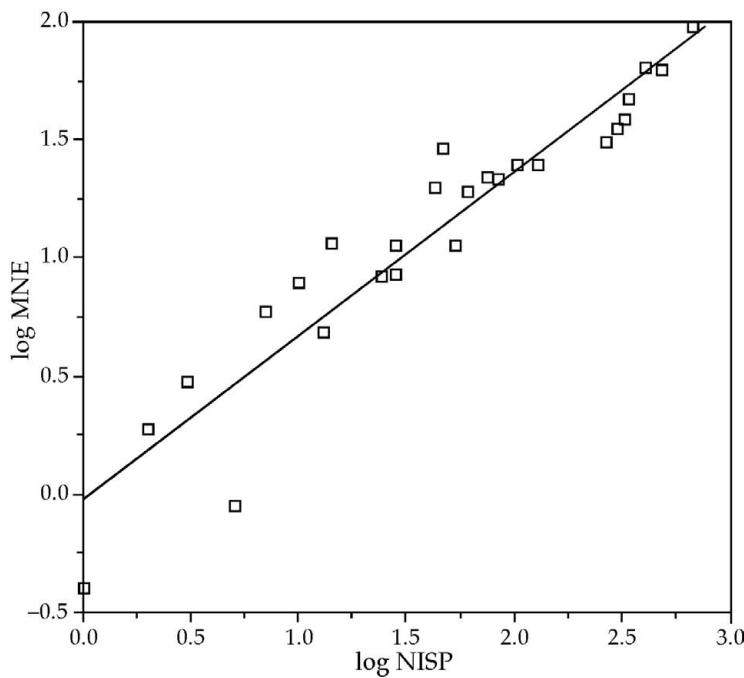


FIGURE 6.18. Relationship between NISP and MNE values for size class II cervids and bovids at Kobeh Cave, Iran. Best-fit regression line ( $Y = -0.015X^{0.692}$ ;  $r = 0.945$ ) is significant ( $p = 0.0001$ ). Data from Table 6.14.

Table 6.15. NISP and MNE frequencies of skeletal parts of saiga antelope (*Saiga tatarica*) from Prolom II Cave, Ukraine. Data from Enloe et al. (2000)

Skeletal part/portion	NISP	MNE
Maxilla	319	81
Mandible	477	56
Sacrum	3	3
Scapula	13	12
Humerus	33	22
Radius	112	46
Carpal	156	48
Metacarpal	138	82
Femur	22	16
Patella	4	4
Tibia	33	26
Lateral malleolus	9	9
Astragalus	82	81
Calcaneum	83	55
Naviculo cuboid	36	36
Cuneiform	18	17
Metatarsal	62	37
First phalanx	285	253
Second phalanx	114	109
Third phalanx	81	72

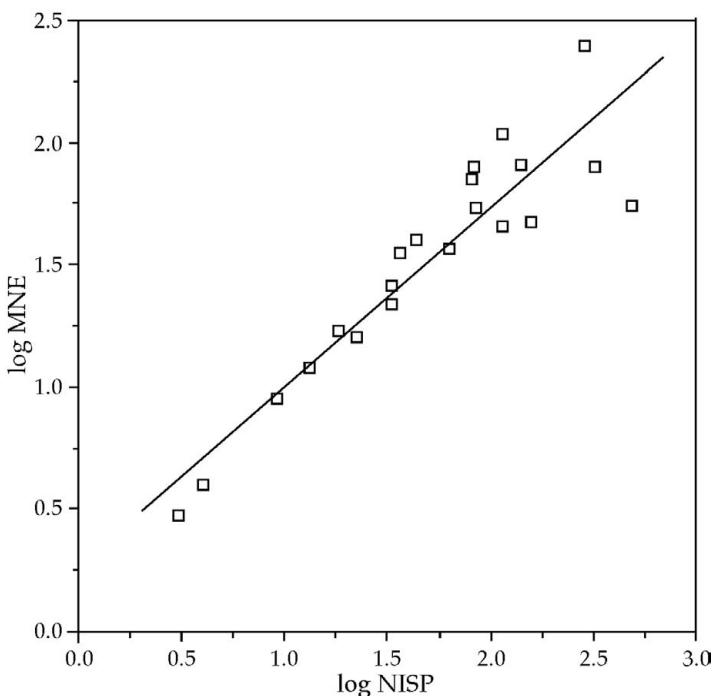


FIGURE 6.19. Relationship between NISP and MNE values for saiga antelope at Prolom II Cave, Ukraine. Best-fit regression line ( $Y = 0.27X^{0.733}$ ;  $r = 0.922$ ) is significant ( $p = 0.0001$ ). Data from Table 6.15.

respect to estimating the frequencies of skeletal parts (Figure 6.19). This is a common pattern. Table 6.16 summarizes the statistical relationships between NISP and MNE for twenty-nine samples of faunal remains. In all but one case,  $r > 0.7$ , and  $p < 0.0001$ ; in twenty-five of twenty-nine cases,  $r > 0.8$ . The value of MNE as a quantitative unit, even though it is explicitly designed to tally frequencies of skeletal parts, seems redundant with NISP; MNE values for an assemblage can often be closely predicted from the NISP values for that assemblage. In fact, Broughton et al. (2006) found the correlation between the two variables to be perfect ( $r = 1.0$ ,  $p < 0.001$ ) in a collection of fish remains accumulated and deposited by an owl.

Earlier in this chapter it was argued that MNE is at best ordinal scale. This is easily shown if we consider the relationship of MNE per skeletal part (or portion) to MNI per skeletal part (or portion). Data revealing this relationship are, like those revealing the relationship between NISP and MNE, also relatively common. We need only graph one data set to illustrate the relationship. Marshall and Pilgram's (1991; Pilgram and Marshall 1995) NISP and MNI data for caprine (*Ovis*

Table 6.16. Relationship between NISP and MNE in twenty-nine assemblages

Assemblage	Relationship	<i>r</i>	<i>p</i>	Reference
Meier deer	$Y = 0.126X^{0.807}$	0.837	0.0001	This volume
Meier wapiti	$Y = 0.063X^{0.766}$	0.883	0.0001	This volume
Kobeh Cave	$Y = -0.015X^{0.692}$	0.945	0.0001	Marean and Kim (1998)
Prolom II Cave	$Y = 0.27X^{0.733}$	0.922	0.0001	Enloe et al. (2000)
Garnsey (bison)	$Y = 0.192X^{0.825}$	0.942	0.0001	Speth (1983)
Sjovold (bison)	$Y = 0.162X^{0.738}$	0.939	0.0001	Dyck and Morlan (1995)
Twilight Cave BS1 – size II	$Y = 0.012X^{0.702}$	0.898	0.0001	Marean (1992)
Twilight Cave RBL2.1 – size I	$Y = 0.029X^{0.786}$	0.956	0.0001	Marean (1992)
Twilight Cave RBL2.1 – size II	$Y = 0.081X^{0.787}$	0.950	0.0001	Marean (1992)
Twilight Cave RBL2.2 – size I	$Y = 0.056X^{0.79}$	0.948	0.0001	Marean (1992)
Twilight Cave RBL2.2 – size II	$Y = 0.063X^{0.76}$	0.937	0.0001	Marean (1992)
Twilight Cave RBL2.3 – size I	$Y = 0.067X^{0.754}$	0.914	0.0001	Marean (1992)
Twilight Cave RBL2.3 – size II	$Y = 0.041X^{0.739}$	0.923	0.0001	Marean (1992)
Twilight Cave DBS – size I	$Y = 0.088X^{0.772}$	0.945	0.0001	Marean (1992)
Twilight Cave DBS – size II	$Y = 0.108X^{0.741}$	0.932	0.0001	Marean (1992)
Friesenhahn Cave ( <i>Homotherium</i> )	$Y = 0.036X^{0.932}$	0.957	0.0001	Marean and Ehrhardt (1995)
Friesenhahn Cave (proboscidean)	$Y = -0.010X^{0.913}$	0.946	0.0001	Marean and Ehrhardt (1995)
Kua Base Camp-size I	$Y = 0.001X^{0.756}$	0.792	0.0001	Bartram and Marean (1999)
Kua Base Camp-size III	$Y = 0.095X^{0.598}$	0.725	0.0001	Bartram and Marean (1999)
Kua Scavenged Kill-size III	$Y = 0.193X^{0.435}$	0.661	0.0001	Bartram and Marean (1999)
Die Kelders-L. 10, size 1	$Y = 0.02X^{0.519}$	0.776	0.0001	Marean et al. (2000)
Die Kelders-L. 10, size 2	$Y = -0.084X^{0.577}$	0.789	0.0001	Marean et al. (2000)
Die Kelders-L. 10, size 3	$Y = -0.173X^{0.592}$	0.732	0.0001	Marean et al. (2000)
Die Kelders-L. 10, size 4	$Y = -0.075X^{0.542}$	0.835	0.0001	Marean et al. (2000)
Nahal Hadera V	$Y = -0.914X^{1.256}$	0.911	0.0001	Munro and Bar-Oz (2005)
Hefzibah	$Y = -0.921X^{1.292}$	0.913	0.0001	Munro and Bar-Oz (2005)
Hayonim Cave-Early Natufian	$Y = 0.046X^{0.763}$	0.815	0.0001	Munro and Bar-Oz (2005)
Hayonim Cave-Late Natufian	$Y = -0.004X^{0.781}$	0.848	0.0001	Munro and Bar-Oz (2005)
el-Wad Terrace	$Y = -0.358X^{1.008}$	0.823	0.0001	Munro and Bar-Oz (2005)

and *Capra*) remains from Ngamuriak, a Neolithic pastoral site in Kenya (Table 6.17), are strongly related (Figure 6.20). Other assemblages from other places and times display the same relationship between NISP and MNI per skeletal part or portion as the Ngamuriak collection. Of the twenty-two assemblages listed in Table 6.18, the relationship between NISP and MNI per skeletal part is rather strong ( $r > 0.7$ ) in twenty assemblages, and it is quite strong in fifteen assemblages ( $r > 0.8$ ). Again, this should come as no surprise given previous discussions and analyses presented in this volume.

Table 6.17. NISP and MNI frequencies of skeletal parts of caprines (*Ovis* and *Capra*) from Ngamuriak, Kenya. P, proximal; D, distal. Data from Pilgram and Marshall (1995)

Skeletal part/portion	NISP	MNI
Tooth rows, lower	288	54
Innominate	133	28
Scapula	94	32
P humerus	30	11
D humerus	91	31
P radius	81	23
D radius	35	16
P metacarpal	50.5	12
D metacarpal	32.5	2
P femur	54	24
D femur	47	12
P tibia	18	6
D tibia	32	12
Calcaneum	65	17
P metatarsal	45.5	12
D metatarsal	29.5	1
First phalanx	83	4
Second phalanx	30	2

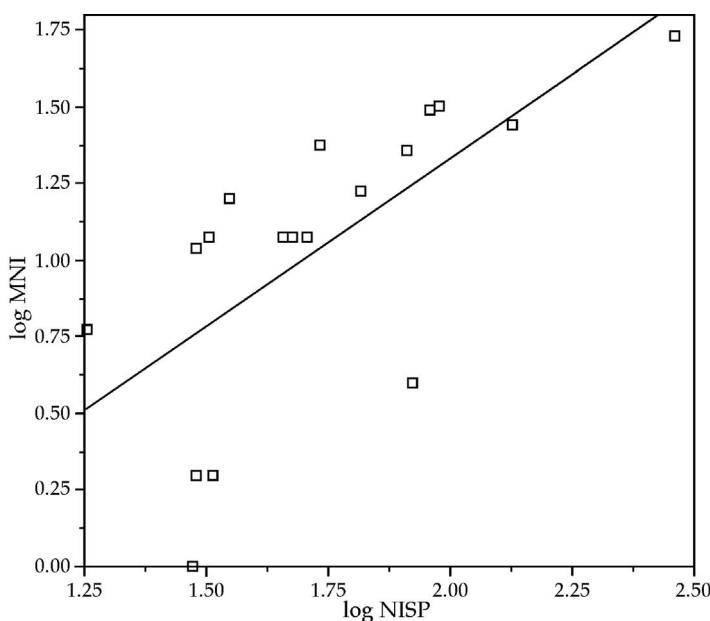


FIGURE 6.20. Relationship between NISP and MNI values for caprine remains from Neolithic pastoral site of Ngamuriak, Kenya. Best-fit regression line ( $Y = -0.469X^{0.904}$ ;  $r = 0.666$ ) is significant ( $p = 0.002$ ). Data from Table 6.17.

Table 6.18. Relationship between NISP and MNI per skeletal part or portion in twenty-two assemblages

Assemblage	Relationship	<i>r</i>	<i>p</i>	Reference
Ngamuriak	$Y = -0.469X^{0.904}$	0.666	0.0026	Pilgram and Marshall (1995)
Gatecliff Shelter	$Y = -0.083X^{0.762}$	0.627	0.0001	Thomas and Mayer (1983)
Boomplaas-Size I	$Y = -0.018X^{0.587}$	0.769	0.0001	Klein and Cruz-Uribe (1984)
Boomplaas-Size IV	$Y = -0.051X^{0.53}$	0.888	0.0001	Klein and Cruz-Uribe (1984)
El Jujo Red Deer-L. 4	$Y = -0.034X^{0.492}$	0.768	0.0001	Klein and Cruz-Uribe (1984)
El Jujo Red Deer-L. 6	$Y = -0.062X^{0.60}$	0.701	0.0001	Klein and Cruz-Uribe (1984)
Equus Cave-Size IV	$Y = -0.083X^{0.705}$	0.888	0.0001	Klein and Cruz-Uribe (1984)
Elandsfontein-Size I	$Y = -0.097X^{0.70}$	0.874	0.0001	Klein and Cruz-Uribe (1991)
Elandsfontein-Size II	$Y = -0.168X^{0.799}$	0.876	0.0001	Klein and Cruz-Uribe (1991)
Elandsfontein-Size III	$Y = -0.395X^{0.977}$	0.905	0.0001	Klein and Cruz-Uribe (1991)
Elandsfontein-Size III	$Y = -0.242X^{0.882}$	0.902	0.0001	Klein and Cruz-Uribe (1991)
Elandsfontein-Size V	$Y = -0.189X^{0.84}$	0.855	0.0001	Klein and Cruz-Uribe (1991)
Klasies River Mouth-Size I	$Y = -0.017X^{0.728}$	0.860	0.0001	Klein (1989)
Klasies River Mouth-Size II	$Y = -0.073X^{0.749}$	0.827	0.0001	Klein (1989)
Klasies River Mouth-Size III	$Y = -0.065X^{0.633}$	0.823	0.0001	Klein (1989)
Klasies River Mouth-Size IV	$Y = -0.014X^{0.684}$	0.748	0.0001	Klein (1989)
Klasies River Mouth-Size V	$Y = -0.017X^{0.582}$	0.779	0.0001	Klein (1989)
El Castillo Cave-Mag & Sol	$Y = -0.102X^{0.877}$	0.958	0.0001	Klein and Cruz-Uribe (1994)
39FA82	$Y = -0.018X^{0.85}$	0.988	0.0001	White (1952)
Bull Pasture bison	$Y = -0.062X^{0.842}$	0.925	0.0001	White (1955)
Bull Pasture wapiti	$Y = -0.074X^{0.884}$	0.966	0.0001	White (1955)
Buffalo Pasture	$Y = -0.152X^{0.901}$	0.883	0.0001	White (1956)

## CONCLUSION

This chapter has concerned the MNE quantitative unit (and various units derived from MNE) and properties it is thought to measure (e.g., skeletal-part abundances, skeletal completeness, fragmentation). As Ringrose (1993:129) pointed out, MNE is a quantitative unit “specifically designed for the study of skeletal-part representation, rather than taxonomic abundance.” This does not mean that MNE and MNI (or NISP) are not mechanically or statistically related. MNI values are by definition (Table 2.4) based on the maximum MNE. Thus, White’s (Table 6.5) summed left and right MNE values are strongly correlated with the MNI values for each of those fourteen skeletal elements (Figure 6.21). This is because MNI per skeletal element is merely the greater of the tally of left elements or the tally of right elements. MNI

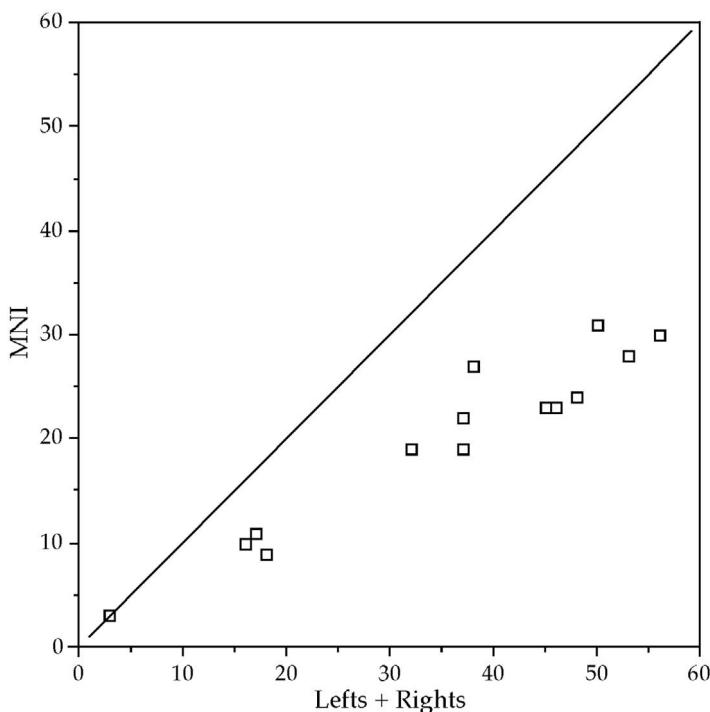


FIGURE 6.21. Relationship between  $\sum(\text{lefts} + \text{rights})$  and MNI per skeletal part. Diagonal shown for reference. Data from Table 6.5.

is a tally of redundant skeletal parts or portions, traditionally based on the greatest MNE in a collection.

MNE seems, on the surface, to be a valuable quantitative unit. It may in fact be valuable if it is clear that, say, femora are much more intensively and extensively fragmented than are humeri. The quick way to determine this is to calculate the relationship between NISP per skeletal part and the MNE per skeletal part. If the two values are strongly correlated, there is little reason statistically to use MNE in further analyses, such as determining if femora are more abundant than humeri; NISP will provide the same ordinal scale information as MNE. MNE is a valuable unit for measuring the intensity of fragmentation, defined as the  $\text{NISP}_i:\text{MNE}_i$  ratio, where  $i$  is a particular skeletal part. Based on analyses and arguments presented in this chapter and elsewhere (Grayson and Frey 2004), MNE is not useful for measuring skeletal-part frequencies. This is so because it is derived (definition dependent), it is influenced by sample size (NISP), and it is influenced by aggregation.

MNE has undergone a history similar to that of MNI. Both units were used to measure the value of a variable, then various potential problems with them were identified

and efforts were made to resolve those problems, for example, tallying units more carefully and more consistently taking into account numerous factors (age/sex/size differences; fragmentation differences and anatomical overlap). After various sorts of potential additional steps to tallying specimens into MNI or MNE units were identified and implemented, it was pointed out that perhaps the quantitative unit is not salvageable despite various safeguards. The quantitative unit is not salvageable because it is in fact a derived measure and it is at best ordinal scale; it is redundant with NISP, a fundamental measure. As with MNI, it seems we have reached the point with MNE where it may no longer be worth using it to the same degree that it once was, particularly with respect to measuring skeletal-part frequencies.

There are other quantitative units similar to MNE. These include the minimum number of butchering units (Lyman 1979; Schulz and Gust 1983), and the minimum number of analytically specified anatomical regions (Stiner 1991, 2002). It is beyond the scope of this discussion to explore the properties of these units, but it is logical to suspect that they, too, will often be strongly correlated with NISP or sample size, and heavily influenced by aggregation and how they are defined. This suspicion is based on the fact that both the minimum number of butchering units and the minimum number of anatomical regions are determined in the same manner as MNE and MNI. The only difference is that the minimum number of butchering units and the minimum number of anatomical regions are at skeletal scales of inclusiveness between MNE and MNI as typically defined. The most important thing to remember is that MNE and similar units are often significantly influenced by sample size, aggregation, and definition, just as is MNI. This simple fact suggests that NISP is to be preferred over MNE and similar units, especially when MNE provides abundance information that is redundant with NISP.

## Tallying for Taphonomy: Weathering, Burning, Corrosion, and Butchering

Taphonomy is a term coined by Russian paleontologist I. A. Efremov (1940) from the Greek words *taphos* (burial) and *nomos* (law). Efremov meant for *taphonomy* to specify the transition, in all details, of organics from the biosphere to the lithosphere. In the context of this book (recall Figure 2.1), taphonomy concerns the agents and process(es) that influence an animal carcass from the moment of that animal's death until its remains (if any survive the vicissitudes of time) are recovered by the paleo-zoologist, and also the kind and magnitude of those influences. There are a plethora of taphonomic agents and processes that variously disarticulate, disperse, alter, and destroy carcass tissues, including bones and teeth (Lyman 1994c).

In this chapter, techniques for tallying what are sometimes referred to as taphonomic signatures, features, or attributes evident on faunal remains are introduced. Identifying the taphonomic agents and processes that influenced an assemblage of faunal remains assists interpretation of the remains. (If the agent is biological, then the taphonomic feature is a *trace fossil* [Gautier 1993; Kowalewski 2002].) Do, for example, those remains reflect what human hunters ate or do they represent a fluvi ally winnowed set of skeletons of animals that died during a seasonal crossing of a river at flood stage? Determination of the taphonomic history of a collection of faunal remains may reveal aspects of paleoecology not otherwise evident among the collection of remains, such as evidence of carnivore gnawing on ungulate bones when no carnivore remains are recovered.

A taphonomic *signature* is a modification feature evident on a skeletal part that is known (or believed) to have been created by only one process or agent (Blumenschine et al. 1996; Fisher 1995; Gifford-Gonzalez 1991; Marean 1995). It is a *signature* because it is unique to that agent or process. A taphonomic feature need not be a signature; it is an artifact or epiphenomenon of an agent or process that modified a skeletal specimen's location, anatomical completeness, or appearance. Given the model of an unmodified skeletal part as it would appear in a normal organism walking, flying,

or swimming around the landscape, any perimortem or postmortem modification to that skeletal part that was not created by physiological processes of the organism (e.g., a healed fracture [antemortem]) is a taphonomic feature. Instances of the occurrence of that modification may be recorded during study of the remains because by definition such an attribute is not a normal feature of a bone or tooth. A modification feature need not have a specifically identifiable cause or creation agent, and in fact many features do not, though that number is decreasing as our knowledge of causal agents increases through actualistic research (e.g., Domínguez-Rodrigo and Barba 2006; Kowalewski 2002). A taphonomic feature is created perimortem (at death) or postmortem (after death). Tallying up, say, the frequency of specimens with gnawing marks, or the frequency of gnawing marks, or both comprises tallying for taphonomy.

Gnawing marks created by hungry carnivores, butchering marks created by hungry hominids, burning damage created by fuel-hungry flames, and various other such taphonomic features can be tallied in various ways to decipher the taphonomic history of a collection of animal remains. Intuitively, for instance, given two collections of bones that are otherwise quite similar (in terms of taxonomic abundances, however measured, and in terms of frequencies of skeletal parts), the one with more bones displaying carnivore gnawing damage is likely the one that underwent relatively more carnivore-gnawing related attrition (consumption) of bone tissue. Tallying such attributes may seem straightforward, but even if tallying is sometimes easy to do, it is not always easy to understand or interpret the tallies. What a tally signifies may well be obscure because a tally of taphonomic attributes (measured variable) may have an unknown relationship to a particular taphonomic (target variable) agent or process. How, for example, does the frequency of gnawed bones (measured variable A) or the frequency of gnawing marks (measured variable B) relate to gnawing intensity (target variable)? Many attributes are thus not *signatures* but are tallied in hope that the quantitative data will reveal aspects of the taphonomic history of the collection.

This chapter begins with some rather easily tallied taphonomic attributes that many taphonomists and paleozoologists believe have well-understood relationships to taphonomic agents and processes. The discussion progresses to complex attributes for which little consensus exists as to how they should be tallied and/or what a tally might mean with respect to a taphonomic agent or process. The goal of this chapter is not to solve particular substantive problems, but rather to describe quantitative units, illustrate how they are tallied, and exemplify how they might be analyzed. And given the topic of this chapter, another quantitative unit must first be introduced.

## YET ANOTHER QUANTITATIVE UNIT

There are two units not often mentioned in the quantitative paleozoology literature that need to be identified. One unit previously mentioned in this book is the number of specimens (NSP). NSP is the number of all specimens in an assemblage or collection (however defined), including those that are identifiable to taxon and those that are not identifiable. NSP is a fundamental measure just like NISP. NSP has been used by name by several zooarchaeologists (e.g., Grayson 1991a; Stiner 2005). Another unit, not previously identified in this book, used by fewer individuals is what Stiner (2005:235) uniquely refers to as the number of unidentified specimens, or NUSP. For any collection of faunal remains,  $NSP = NISP + NUSP$ , and  $NISP = NSP - NUSP$ .

Why should NSP and NUSP be of concern? First, and less importantly, they need to be mentioned because sometimes paleozoologists will refer to the ratio NISP/NSP. The implication of the ratio is seldom stated explicitly, but it seems to be thought that the higher the ratio (the greater the proportional value of NISP), the more specimens were identified because they were not so badly preserved (corroded, fragmented) as to be unidentifiable. The NISP/NSP ratio is thus thought to reflect general aspects of the taphonomic history of a collection (e.g., fragmentation and destruction extent and intensity). Perhaps because the relationship of the NISP/NSP ratio to preservational condition has never been empirically or critically examined, the NISP/NSP ratio is seldom used analytically. Or, perhaps the ratio is seldom analyzed because it could be a function of which skeletal parts are represented as some parts are more easily identified than others. Whatever the case, if mentioned at all, the ratio is often mentioned in a descriptive role. After all, it is simple to calculate and it is based on two directly measured variables – NISP and NSP – that are readily determined. (NSP and NUSP are influenced by fragmentation, though this is not generally acknowledged.)

The second reason to mention NSP and NUSP is important and concerns the fact that many of the features tallied for taphonomic purposes can be tallied for the NISP of a collection, or for the NSP (= NISP + NUSP). Here taphonomic features are discussed as if they are only tallied using NISP because that is the traditional manner in which they are tallied. Distinguishing NISP and NUSP, and tallying the taphonomic features using both, may be worth considering *if*, and this is important, there is reason to believe that the taphonomic process or agent might be reflected differently across identified specimens than it is across unidentified specimens. If there is no reason to believe that this is the case, then if the sample of NISP is sufficiently large (and we can ascertain that by sampling to redundancy [Chapter 4]), then tallying taphonomic features across NISP will likely be sufficient to measure taphonomic

Table 7.1. *Weathering stages as defined by Behrensmeyer (1978)*

Stage	Definition
0	Greasy, no cracking or flaking, may have soft tissue attached.
1	Longitudinal cracking; articular surfaces with mosaic cracking; split lines beginning to form.
2	Flaking of outer surface (exfoliation); cracks present; crack edges are angular.
3	Compact bone has rough, fibrous texture; weathering penetrates 1–1.5 mm; cracked edges are rounded.
4	Coarsely fibrous and rough surface; loose splinters present; weathering penetrates to inner cavities; cracks are open.
5	Bone tissue very fragile and falling apart; large splinters present.

variables and the degree or extent of the influence of the taphonomic processes and agents of concern.

## WEATHERING

In a classic paper, Behrensmeyer (1978) specified six stages through which a (mammal) bone would pass during subaerial weathering (Table 7.1). She defined weathering as “the process by which the original microscopic organic and inorganic components of bone are separated from each other and destroyed by physical and chemical agents operating on the bone *in situ*, either on the [Earth’s] surface or within [sediments]” (Behrensmeyer 1978:153). Weathering involves the natural decomposition and destruction of bone tissue, and Behrensmeyer recorded subaerial weathering only – weathering that occurs “on the [ground] surface.” To quantify weathering damage, Behrensmeyer (1978:152) suggested that the analyst tally the number of bone specimens that display each weathering stage. The weathering stage that a specimen displays is, in turn, recorded as the maximum weathering stage evident on an area comprising at least 1 cm<sup>2</sup> of the surface of a specimen.

The maximum weathering displayed is used because the target variable of interest concerns not how weathered (or unweathered) a specimen is, but rather the duration of “surface exposure of a bone prior to burial and the time period over which bones accumulated” (Behrensmeyer 1978:161). This means that the maximum weathering stage evident is recorded rather than minimum or average weathering evident on a specimen for two reasons. First, several variables mediate (slow) the rate of weathering (Lyman and Fox 1989), thereby potentially weakening any statistical relationship between weathering stage (the dependent variable) and the variable of analytical

Table 7.2. Weathering stage data for two collections of mammal remains from Olduvai Gorge. Frequencies are NISP (% of total NISP). Data from Potts (1986)

Stage	FLK "Zinj"	FLKNN L/2
0	771 (76)	105 (46)
1	147 (14)	59 (26)
2	63 (6)	24 (10)
3	36 (4)	39 (17)
4	0 (0)	2 (1)
5	1 (0)	1 (0)

interest (e.g., duration of exposure). The second reason that maximum weathering is used is that there is a strong statistical relationship between date of death of the animal contributing the maximally weathered bone and the maximum weathering stage displayed by one or more bones of the carcass (Behrensmeyer 1978). The critical interpretive issue, then, requires understanding the relationship between maximum weathering stage displayed and the target variable of interest whether it be duration of exposure, date of animal death, or something else.

Quantifying bone weathering is relatively straightforward. Count up how many specimens in a collection display each of the six weathering stages. Then, present the tallies in a table as absolute counts of specimens per weathering stage, as proportions or percentages of specimens per weathering stage, or both. Data can also be presented graphically. An example is provided by Potts's (1986) data for assemblages of mammal remains from Plio-Pleistocene archaeological sites in Olduvai Gorge, Tanzania. Data for two assemblages are summarized in Table 7.2, and percentage frequency data for the two assemblages are graphed in Figure 7.1 in what Lyman and Fox (1989:300) term a *weathering profile*, defined as "the percentage frequencies of bone specimens in an assemblage displaying each weathering stage." Percentage frequency data eliminate the effects of variation in sample size, thereby permitting differences between the two assemblages plotted in the graph in Figure 7.1 to be interpreted in terms of differences in weathering rather than in terms of difference in sample size. The fact that one assemblage is nearly four and a half times larger than the other (Table 7.2) is not apparent in Figure 7.1.

Note that thus far the weathering data have been presented in tabular form (Table 7.2) and in graphic form (Figure 7.1). How might those data be analyzed further?  $\chi^2$  analysis indicates that specimens are not equally distributed within the weathering stages across the two assemblages ( $\chi^2 = 109.74$ ,  $p < 0.0001$ ). Analysis

Table 7.3. *Expected frequencies (EXP) of specimens per weathering stage (WS) in two collections (Zinj; L/2), adjusted residuals (AR), and probability values (p) for each.* Based on data in Table 7.2

WS	Zinj EXP	L/2 EXP	Zinj AR	L/2 AR	Zinj p	L/2 p
0	714.6	161.4	9.09	-8.95	<0.01	<0.01
1	168.0	38.0	-4.19	4.11	<0.01	<0.01
2	71.0	16.0	-2.31	2.30	<0.05	<0.05
3	61.2	13.8	-7.84	7.77	<0.01	<0.01
4	1.6	0.4	-2.90	2.95	<0.01	<0.01
5	1.6	0.4	-1.09	1.10	>0.1	>0.1

of adjusted residuals indicates that relative to the FLKNN L/2 assemblage, the FLK “Zinj” assemblage contains more specimens displaying weathering stage 0 and fewer displaying stages 1, 2, 3, and 4 than expected given random chance (Table 7.3). These results identify the statistical significance of Figure 7.1, but there is other variation between the two weathering profiles that the  $\chi^2$  analysis does not capture. What other kinds of analysis might be done?

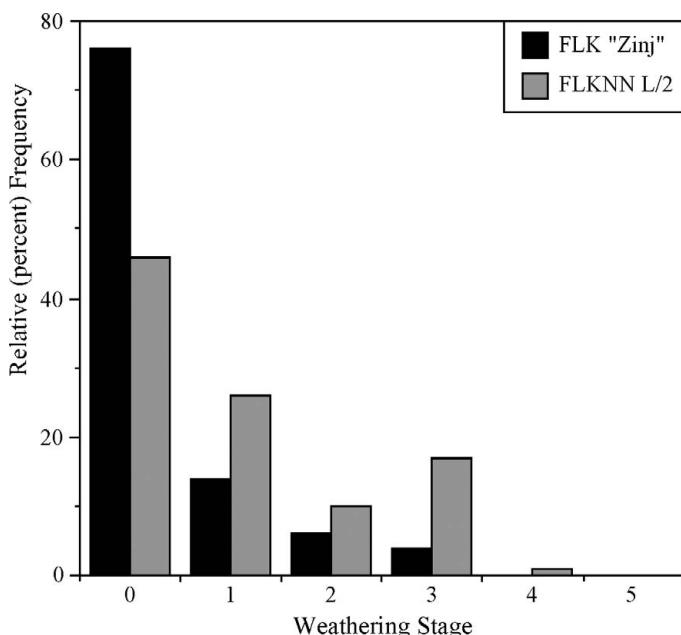


FIGURE 7.1. Weathering profiles for two collections of ungulate remains from Olduvai Gorge. Data from Table 7.2.

A paleozoologist could also determine the richness of weathering stages in each, and the evenness and heterogeneity of each as well. These values for FLK “Zinj” are 5 (richness), 0.489 (evenness), and 0.787 (heterogeneity); for FLKNN L/2 the values are 6, 0.730 and 1.309. The richness values do not tell us much by themselves. The evenness value is greater for FLKNN L/2, indicating that it has a more even distribution of specimens across the weathering stages than the FLK “Zinj” collection. Finally, the heterogeneity index values conform with the combined richness and evenness values for each and indicate that the FLKNN L/2 assemblage is more heterogeneous – richer and more even – than the FLK “Zinj” assemblage. All of these values, and particularly the evenness and heterogeneity index values, suggest that the FLKNN L/2 assemblage is more weathered than the FLK “Zinj” assemblage, although the only way we know this rather than it being the other way around is in light of Figure 7.1.

Tallying so far has been straightforward. But if one were to interpret the data in Tables 7.2 and 7.3 and the graph in Figure 7.1 as reflecting differences in bone accumulation duration, as Potts (1986) did and Behrensmeyer (1978) hoped to do – the basic presumption being that an assemblage with a more left-skewed weathering profile (tail to the left, maximum frequency to the right) took longer to accumulate than an assemblage with a right skewed profile (tail to the right, maximum frequency to the left) – a number of assumptions would have to be made. Behrensmeyer (1978) perceived a positive relationship between how long a carcass had been lying on the landscape (years since death) and the greatest weathering stage displayed by any one of the bones of the carcass. The correlation between the two variables is indeed strong and significant, as implied by Figure 7.2 ( $r = 0.872$ ,  $p < 0.0001$ ). Based on actualistic research by numerous others, Lyman and Fox (1989) noted that there were a number of variables that could reduce the correlation coefficient to considerably less than 1.0. Few individuals subsequently presented, let alone interpreted, bone weathering data as Potts (1986) had done. Rather, weathering data came to be used to gain insight into other aspects of the taphonomic history of a bone assemblage.

Given Behrensmeyer’s (1978:153) observation that “bones are usually weathered more on the upper (exposed) than the lower (ground contact) surfaces,” analysts now examine which surface is more weathered and which is less weathered. Skyward or upper surfaces *should* be more weathered than groundward or lower surfaces because the upward surface is more directly exposed to sunlight, precipitation, and other climatically related weathering agents. If the reverse is observed, if the groundward surface is more weathered than the upper surface, then it is likely there has been some postdepositional and perhaps postburial disturbance. If there are many bones, then one could construct a weathering profile like that in Figure 7.1, but with a distinction between the weathering stage displayed by the skyward surfaces and the

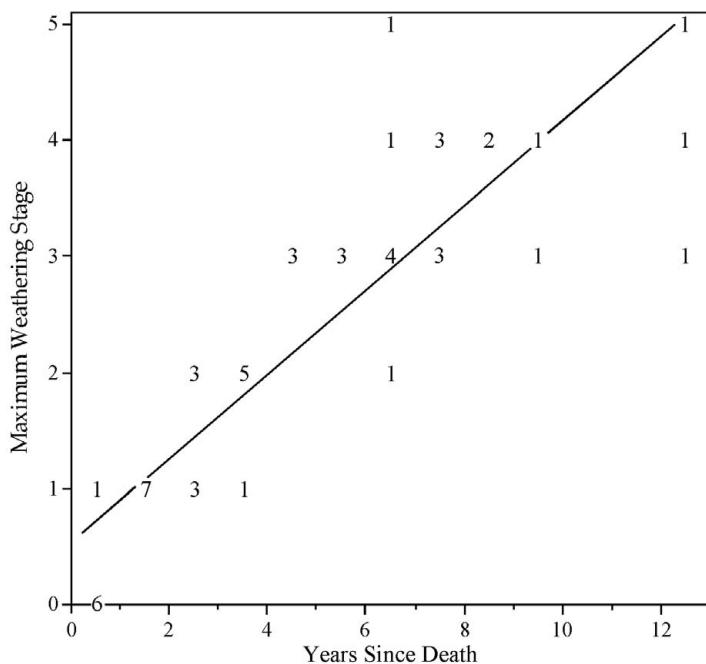


FIGURE 7.2. Relationship between years since death and the maximum weathering stage displayed by bones of a carcass ( $r = 0.872$ ,  $p < 0.0001$ ). Plotted numbers indicate frequency of carcasses displaying a particular weathering stage and years since death. Data from Behrensmeyer (1978).

weathering stage displayed by the groundward surfaces. An example of such a graph using fictional data is shown in Figure 7.3. What is shown is what is expected in an assemblage that experienced minimal post depositional disturbance – the groundward surfaces are generally less weathered than the skyward surfaces.

No taphonomist has used a quantitative unit for tallying weathering data other than NISP. Some analysts, beginning with Behrensmeyer (1978), have suggested that perhaps frequencies of weathered long bones should be tallied separately from frequencies of small, compact bones, such as carpals, tarsals, and phalanges, and perhaps as well scapula, innomates, skulls, and vertebrae should be tallied as a group separate from long bones (one group) and small bones (another group). Do femora weather at the same rate as phalanges? Actualistic data suggest that they do not (summarized in Lyman and Fox 1989; see also Lyman 1994c). Thus, one might tally the number of each skeletal element displaying each weathering stage. If comparison of the weathering profiles of, say, scapulae and humeri are not significantly different, then lump them together for a skeletal composite weathering profile. Multiple such comparisons between different categories of bone size and shape may

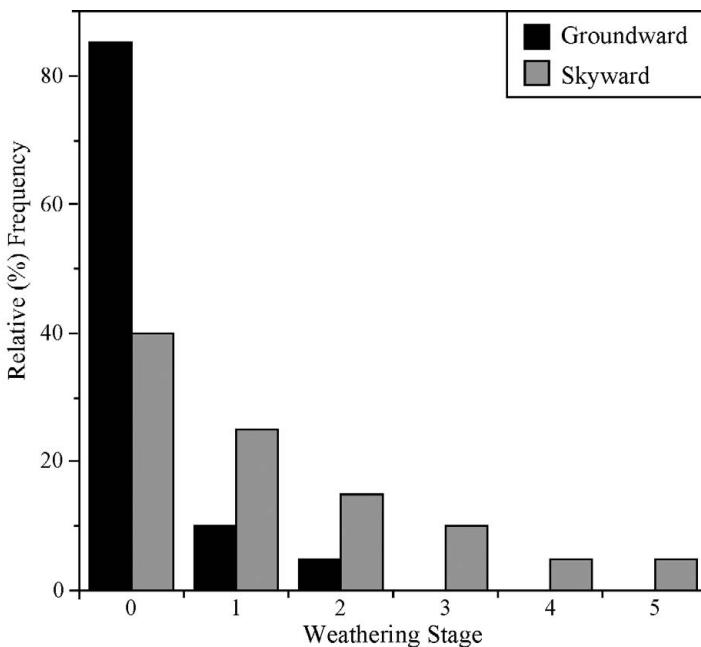


FIGURE 7.3. Weathering profiles based on fictional data for a collection of bones with skyward surfaces representing one profile and groundward surfaces representing another profile.

indicate that, say, small, more or less spherically shaped dense bones such as carpal, tarsals, and phalanges, and larger, plate-shaped less dense bones, such as scapula and innominate, may reveal differences in weathering profiles (for a discussion of how to measure bone shape, see Darwent and Lyman [2002]). Similarly, one might separately tally weathering stages evident on bones of large ungulates and those evident on bones of small ungulates, or equids and bovids, or the like to evaluate similarities or differences between taxa in terms of weathering.

Quantification of bone weathering data involves counting bones or counting kinds of bones (skeletal element, or bone shape) based on the maximum weathering stage displayed by each. The basic counting unit is NISP, but different kinds of NISP (shape, size, taxon considered or not) may be distinguished. The entire specimen (regardless of kind) is subject to weathering yet the longest exposed (if you will) portion of the specimen will display the maximum degree of weathering. Recall why the most advanced weathering staged displayed by a specimen is recorded rather than a less advanced stage. The conceptual clarity of the relationship between the target variable and the measured variable is what makes the quantification of bone weathering attributes straightforward relative to some other kinds of taphonomic

attributes. The entire surface of each discrete specimen can potentially weather at the same rate – all surfaces are taphonomically interdependent at a general level because they are all connected – and although this does not seem to happen in practice (groundward surfaces weather slower than skyward surfaces), interdependence of all surface area of a specimen renders NISP the correct quantitative unit if one wishes to know which of two assemblages of bones is the most weathered.

If post-depositional disturbance is of interest, then knowing the stage of weathering displayed by skyward and by groundward surfaces is important. If intraskeletal variation in weathering is of interest, then tally by distinct skeletal parts. If taxonomic variation in weathering is of interest, then tally skeletal parts by taxon. The target variable and how to tally should be specified by the research question. In most cases, each specimen will be tallied based on the maximum weathering it displays. Weathering stage data are generally tallied using NISP. One might tally weathering by NSP to determine if a higher proportion of NUSP displays more advanced weathering than NISP; if so, that would suggest long-term exposure on the ground surface and a low NISP:NSP ratio that resulted from subaerial weathering. The interdependence of the entire surface of a specimen dictates the tallying protocol, and it also attends the tallying of other sorts of taphonomic modifications to bones.

#### CHEMICAL CORROSION AND MECHANICAL ABRASION

Some faunal remains have passed through a digestive tract and as a result have been chemically corroded (e.g., Andrews 1990). Corrosion features on bones include solution pits or ovoid depressions, fissures that penetrate through cortical bone, and feathered fracture surfaces (Darwent and Lyman 2002; Klippel et al. 1987; Lyman 1994c). Quantification of such observations typically involves tallying the NISP that display digestive (or other) corrosion and then calculating the percentage of the total NISP that display corrosion (e.g., Fernandez-Jalvo and Andrews 1992; Klippel et al. 1987; Weissbrod et al. 2005). (Seldom is the percentage of NSP that displays digestive corrosion reported; it might be taphonomically revealing to compare the percentage of NISP that has been digested with the percentage of NUSP that displays digestive damage.) Because the entire specimen is ingested, all areas of the surface of the specimen are interdependent with respect to the action of the taphonomic process of digestive corrosion. NISP (or NSP) thus is the logical quantitative unit for tallying digestive corrosion.

Given that stages of the degree of corrosion can be defined (e.g., Darwent and Lyman 2002; Fernandez-Jalvo and Andrews 1992; Matthews 2002), one might

construct a *digestive corrosion profile* analogous to a weathering profile (Figure 7.1). The frequency data used to construct the profile would allow graphical and statistical comparisons between assemblages of bones that may have undergone different levels of digestive corrosion like those applied to weathering data. We do not yet know enough about digestion and how, say, hair as in owl pellets might buffer some portions of a bone specimen's surface area from corrosion. Once we know this, some research questions may demand that the range of the degree of corrosion damage to a specimen be recorded.

What is often referred to as root etching is another kind of corrosive damage that is sometimes reported. Again, taphonomic interdependence of all areas of the surface of the specimen makes NISP the quantitative unit of choice. But, upward surfaces may display root etching whereas downward surfaces of specimens do not (this seems to be the case). If so, and as yet we seem to know too little about root etching to be sure, then this sort of corrosive damage could be quantified by number of skyward surfaces and number of groundward surfaces. Until we know more about the taphonomic process itself, quantitative units have ambiguous significance; we do not presently know what corrosion damage quantified using NISP means. Exactly the same can be said for frictional abrasion such as occurs during fluvial transport and other sorts of damage that have the potential to influence the entire surface of a specimen. Until we know better, NISP is the quantitative unit of choice for measuring corrosion and abrasion.

But there may be a better quantitative unit for such features. One might measure the amount (proportion) of surface area that displays corrosion, erosion, and abrasion damage, although this would require a labor-intensive analysis. Whether one uses NISP or the amount of surface area as the quantitative unit will depend on the target variable. Often, what that target variable might be is unclear in the literature. If the desire is to compare two assemblages, then it may make no difference whether percent of surface area or percent of NISP that displays a kind of damage is determined. Until we have better knowledge of the relationship between particular measured variables and particular target variables, using NISP will suffice. The same argument applies with equal force to quantifying the taphonomic signature of the effect of fire on faunal remains.

#### BURNING AND CHARRING

Quantification of burned bone has, like other taphonomic features, typically involved determination of the percentage of the total NISP that comprises burned specimens

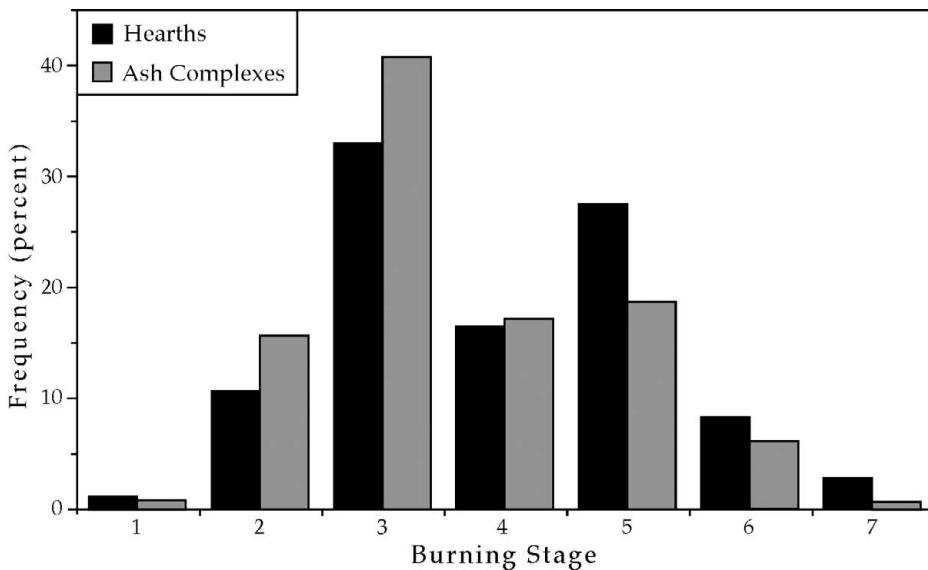


FIGURE 7.4. Frequency distribution of seven classes of burned bones in two kinds of archaeological contexts. Original data from Cain (2005).

(e.g., Cain 2005; Edgar and Sciulli 2006; Grayson 1988; Pozorski 1979). One could also tally the percentage of NSP that has been burned. Because *burning* is a process that results in continuous modification of bone tissue – burning is a result of excessive heat – different stages of burning can often be distinguished. One of the simplest schemes to operationalize was specified by Brain (1981:55) who designated three stages: (1) unburned bone; (2) “carbonized” bone that is black because as the collagen is burned a specimen becomes charcoal or carbonized; and (3) “calcined” bone that is white because continued heating has oxidized the carbon. There is the potential for even finer distinctions of how intensively burned individual specimens are (e.g., Cain 2005; Johnson 1989; Shipman et al. 1984).

Tallying specimens by the maximum burning stage each displays, one can construct a “burning profile” much like a weathering profile. Figure 7.4 shows the percentage frequencies of bone specimens <2 cm long from the Middle Stone Age site of Sibudu Cave, South Africa. Cain (2005) presented these data for six individual hearths and two individual “ash complexes.” His data are summed by functional context (hearth vs. ash complex) in Figure 7.4. That figure suggests burning is similar across the two kinds of contexts. Cain (2005) does not present his data in a manner that allows  $\chi^2$  analysis, analysis of adjusted residuals, or calculation of the evenness of representation of burning stages such as was done with the weathering data in Figure 7.1. These sorts of quantitative analyses would be logical next steps.

It may be informative if bones are tallied in categories distinguished by burning stage and also by taxon and skeletal element represented (or NISP and NUSP, given that more specimens in the more advanced stages of burning in NUSP than in NISP would suggest burning related fragmentation reduced the proportion of identifiable specimens). Thus, say, one could have burning profiles the tally categories of which are identical to those described for weathering. And, like with weathering, record the maximum burning stage represented over an area of at least 1 cm<sup>2</sup>. In the absence of insulating soft tissue, all areas of the surface of each specimen are taphonomically interdependent and unless the specimen is quite large the entire specimen may well undergo the same degree of heating. One could indicate on drawings of bones which portions of each specimen are unburned, carbonized, and calcined, but without a very specific research question or hypothesis that demands such data (an explicit target variable), and a solid (actualistically evaluated) interpretive model for making sense of such data (mechanical linkages between burned and unburned portions of each specimen and the heating regime), recording which surfaces are burned and which are not, irrespective of the boundaries of the discrete specimen seems unnecessary.

One critical point concerns (in)explicit identification of the target variable. Is it merely the proportion (or percent) of burned specimens? Is it the *intensity* of burning? If intensity of burning is the target variable, then the manner in which that variable is defined by the analyst will specify the appropriate variable to measure. If intensity is defined in terms of the *amount* of burned bone, then surface area would be the appropriate measured variable. But if intensity is defined in terms of the number of burned specimens, then it would be better to use NSP (or NISP) to quantify burning, and to determine the percent (or proportion) of specimens that had been burned because that measured variable is the definition of the target variable. Whatever the case, the validity of all of the measured variables for assessing a particular target variable is presently unclear because the nature of the relationship between the two is unknown. Nevertheless, the discussion to this point warrants a brief digression.

### *A Digression*

All of the taphonomic processes that have been discussed thus far create modifications on the surface of a specimen, but all of them may, for myriad reasons, modify only a fraction of the total surface area of a specimen. This fact can be taken advantage of when tallying for taphonomy. If the portions of the total surface area of each specimen are not interdependent with respect to the operation of a taphonomic

process or agent, then rather than tally the total NISP, the total NISP with the modification of interest, and division of the latter by the former to derive a %NISP with the modification, a different quantitative protocol can be applied – this can be referred to as the *surface area solution*.

The ARC GIS procedure described in Chapter 6 and first described by Marean et al. (2001) might prove to be a good means for measuring the amount of total surface area that displays maximum weathering, corrosion, burning, or other taphonomic attributes. If this procedure is chosen, then the amount of surface area measured will depend on how accurately specimens are “mapped” onto the computer-stored templates of skeletal elements, and also how accurately weathered, corroded, and burned areas are mapped onto the templates. More importantly epistemologically, however, is the fact that the taphonomic significance of measurements of amount of weathered, corroded, and burned surface area is unclear. We do not know enough about taphonomic processes to be able to interpret these surface area data with any validity. The surface area solution is thus but one way to measure taphonomic attributes. It may not be worth pursuing because taphonomic processes are so historically contingent as to be beyond the analytical fine-scale resolution provided by measures of proportions of modified surface area (Lyman 1994c, 2004c). Description of other means to quantify other sorts of taphonomic features will clarify this.

#### GNAWING DAMAGE

It has long been known that various animals gnaw bones for diverse reasons (Fisher [1995] and references therein). Generally, damage created by rodent gnawing is readily distinguished from gnawing damage created by carnivores (Fisher 1995; Noe-Nygaard 1989). Taphonomic knowledge is sufficient to allow the distinction of several kinds of damage created by carnivores; these include punctures, furrows, and irregular damage (e.g., Haynes 1980). The NISP (NSP, less often) of gnawed specimens is usually tallied, and the relative (percentage) abundance of gnawed specimens [ $100 \times \text{NISP gnawed}/(\text{NISP gnawed} + \text{NISP not gnawed})$ ] determined for each taxon, for each skeletal part or portion of a taxon, or however the analyst believes the data should be structured (which will depend in part on the research question, which in turn should explicate the target variable and the measured variable) (e.g., Todd et al. 1997). Thus, one could tally the frequency of femora that display punctures and the frequency that display furrows, and compare those to the frequency of humeri that display punctures and the frequency that display furrows, respectively. Or, determine the relative frequency of gnawed femora and the relative frequency of gnawed humeri

to determine if femora were more extensively (or frequently) gnawed than humeri (e.g., Cruz-Uribe and Klein 1994). The percentage of specimens of taxon A that have been gnawed can also be compared with the percentage of specimens of taxon B that have been gnawed (e.g., Cruz-Uribe and Klein 1994).

The frequency of gnawed specimens is often interpreted as a measure of the intensity of gnawing. One might argue, however, that the number of gnawing marks per gnawed specimen is a better measure of gnawing intensity and whether the remains of one taxon or one kind of skeletal element has been more intensively gnawed than the remains of another taxon or another kind of skeletal element, respectively. Such a measure demands two things, one logical and one practical. The logical requirement is that “intensity of gnawing” be clearly defined. The practical requirement is that individual gnawing marks – ones made by each biting action or each instance of dragging teeth across a bone surface – be distinguishable from one another, but they often are not. Perhaps, however, this is not really a problem. Is a gnawing mark a single puncture, or furrow, or instance of irregular damage? Was each individual puncture or furrow or instance of irregular damage created by one bite, or one instance of teeth contacting bone or dragging across the surface of a specimen?

To answer the last two questions requires that we define *intensity* of gnawing. Most taphonomists likely mean how damaged the specimens are or how much energy was expended by the bone gnawer. More gnawing marks may well mean more energy was expended, or it might not (Kent 1981). More gnawing marks may simply mean greater damage to bone surfaces. This ambiguous target variable brings us back to how to tally such damage, regardless of the meaning of that damage or its frequency. Let’s assume we can distinguish individual gnawing marks. Whereas individual punctures and furrows could each be tallied as “1,” the category known as irregular damage presents a problem because individual tooth marks are typically indistinguishable in such damage. (The lack of distinguishability is particularly acute with respect to individual rodent gnawing marks [Thornton and Fee 2001].) But irregular damage also suggests a different way to measure the intensity of gnawing damage.

One could determine the amount of surface area that has been destroyed by gnawing. Given that there is little clear taphonomic significance to the difference between tooth punctures, furrows, and irregular damage, the amount of surface area that has been destroyed by such modifications may well provide a robust measure of the intensity and degree and extent of gnawing damage. And, such a measure does not demand that individual tooth marks be distinguished within an irregularly damaged area. Rather, only the amount of the total surface area of all specimens could be inspected for gnawing damage, and the amount (percentage) of surface area actually damaged, regardless of the type of damage (punctures, furrows, irregular) could be

determined. One could determine such a value for both damage created by rodent gnawing and damage created by carnivore gnawing, if desired. This brings us back to the surface-area solution to tallying for taphonomy when burning, weathering, and corrosion damage were under consideration. If that protocol is used, then the problem reduces to accurate mapping of specimen borders and of boundaries of damaged surfaces (see the discussion in Chapter 6).

Assuming that one can accurately determine the amount of damaged surface area, other questions arise. Should the amount of corroded/abraded surface area be subtracted from the total surface area inspected for gnawing damage? What about the amount of weathered surface area? Answers to these sorts of taphonomic questions need to be in hand before too much energy is spent designing new ways to quantify traces of taphonomic processes and agents. Thus, for the present, the ratio of gnawed specimens to gnawed plus ungnawed specimens, expressed as a percentage or proportion, is an acceptable measure of gnawing intensity. There is a final, general category of damage to bone surfaces that, at least from a zooarchaeological perspective, may help bring the preceding portion of this chapter into fine-resolution focus.

#### BUTCHERING MARKS

*Butchering* is the human reduction and modification of an animal carcass into usable or consumable parts (Lyman 1987a, 1992b, 2005b). It involves the set of hominid behaviors and activities that occur between the time of carcass procurement (regardless of how it is procured [e.g., hunted or scavenged] or its condition) and final disposal or abandonment of variously consumed, and unconsumed, used and unused portions of the carcass. Butchering behaviors occur in varying orders and frequencies or intensities at various times for different carcasses because butchering is historically contingent, which means that the particular order and frequency of individual behaviors depends on a plethora of variables such as carcass size, carcass location on the landscape, time of day, air temperature, number of butchers, and butchering tools available. Butchering activities have traditionally been categorized as belonging to one of three or four basic kinds: skinning, dismemberment or disarticulation, filleting or removing meat from bones, and marrow and grease extraction (Binford 1981; Guilday et al. 1962; Noe-Nygaard 1977; Pozorski 1979). The first two sets of activities are focused on reducing a carcass into manageable pieces whereas the third focuses on extraction of consumable meat external to the bones and the last set of activities focuses on extraction of within-bone nutrients. There are a plethora of other activities involved, including evisceration, extraction of blood, brains, bone,

and sinew, and periosteum removal that can take place but that can be subsumed within one of the three or four traditionally recognized general activities.

Each butchering activity, regardless of how it is categorized, can produce what are typically called *butchering marks*. Many believe that such marks can be reliably and validly identified (e.g., Blumenshine et al. 1996; Fisher 1995), so here the focus is on how these marks are quantified. The reasons to worry about counting butchering marks are several. Most simply, butchering is a process; it begins with a single discrete entity (carcass) and ends up with multiple discrete entities (disarticulated and disassociated complete and incomplete skeletal elements, hide, brain, marrow, muscle masses, etc.). As butchering progresses, the carcass is reduced into successively more numerous discrete pieces. The butchering process often involves the application of various kinds of forces to the carcass to reduce it into consumable and usable pieces. These forces can modify bones by breaking them and they can modify bone surfaces by scarring them. It is the marks that are created by butchering that are of interest here; quantitative measures of fragmentation are discussed in Chapter 6.

As the butchering process continues, more marks may be created on the bones of a carcass. It is likely for this reason that many zooarchaeologists have sought to measure the *intensity* of butchering, by which is meant, it seems, the amount of energy invested in butchering. Butchering intensity is measured by tallying butchering damage evident on a collection of faunal remains (e.g., Haynes 2002). For example, Binford (1988:127) suggested that “the number of cut marks, exclusive of dismemberment marks, is a function of differential investment in meat or tissue removal.” Other zooarchaeologists have tallied frequencies of various kinds of butchering marks for other reasons. Bunn and Kroll (1986:432), for example, state that “frequencies of cut marks on different skeletal parts can be directly linked to the skinning, disarticulation, and defleshing of carcasses” and that multiple occurrences of marks in a particular anatomical location indicate, say, “repeated dismemberment of the elbow joint.” Regardless of the reason for tallying frequencies of butchering damage, if they are to be tallied, we must first have explicit definitions of what the various kinds of damage are. It is to that topic that we turn next.

### *Types of Butchering Damage*

There are several variables that may be considered when tallying butchering marks, but one of them is virtually always considered. That variable concerns the type of mark. There are several basic kinds of marks the morphologies of which are dependent on the type of force and aspects of force application used to create them (Fisher 1995;

Greenfield 1999; Lyman 1987a; Noe-Nygaard 1989; Potter 2005; Thompson 2005). It is likely because of the different kinds of force and different ways that force is applied through an intermediary (tool) to a bone surface that most zooarchaeologists can reliably and validly distinguish the various mark types that have been recognized (Blumenschine et al. 1996).

One kind of force involves dynamic percussion, such as when a hammer stone impacts a bone resting on a firm surface. This type of force application involves a more or less blunt (as opposed to sharp-edged) implement and abrupt dynamic loading (impact) that produces impact notches, flake scars, and various scratches (Blumenschine and Selvaggio 1988; Pickering and Egeland 2006). Another kind of force involves sawing or slicing forces that produce what are termed “cut marks” or “striae.” Sawing and slicing involves force application parallel to the long axis of the cutting edge of the tool. *Scraping* is similar to slicing and sawing, though the latter two are generally back and forth whereas the former is generally in one direction and force is applied perpendicular to the long axis of the implement’s working edge. *Chopping* is dynamic loading with a sharp edge; Gifford-Gonzalez (1989) considers it to be a cutting-like process, and although it can be, I conceive of it as something of a hybrid between cutting and percussion.

Percussion marks, cut marks, and scraping marks tend to have been produced in prehistoric contexts by butchers with primitive (preindustrial) technologies. Chopping marks are made in such contexts as well, but they are also made in historic contexts by butchers with industrial-grade (metal) technologies (Landon 1996). So, too are saw cuts. By the latter is meant cuts made with metal saws (e.g., Lyman 1977). Saw cuts can be tallied in various ways, most of which are the same as the ways used to tally cut marks, percussion marks, and chopping and scraping marks. For the sake of simplicity, discussion is limited to percussion marks and cut marks in the following.

#### *Tallying Butchering Evidence: General Comments*

A classic statement in zooarchaeology is this: “It is quite possible to butcher an animal of any size without leaving a single [butchering] mark on any bone” (Guilday et al. 1962:64). This claim was reiterated at least twice more in later years (Bunn and Kroll 1988; Crader 1983), so it is perhaps not surprising that numerous analysts subsequently suggested various reasons why a bone might not display butchering marks despite the fact that the portion of the carcass represented by that bone had apparently been butchered. Shipman and Rose (1983:86) suggested that “soft tissues have an ability to shield bones from being marked by bone or stone tools.” They found

in an experimental context that even the periosteum (a <1 mm thick, soft-tissue covering of bone) shields bone surfaces from cut marks (Shipman and Rose 1983:70).

Gifford-Gonzalez (1989:202) made a similar observation in an ethnoarchaeological context. Olsen and Shipman (1988:545) argued “butchering requires a light touch to prevent crushing and dulling the tool’s edge by contact with the bone,” and a butcher’s desire to not dull a cutting tool would result in few butchering marks. Guilday et al. (1962:64) thought that the probability that a bone will display a butchering mark is a function of “the skill of the [butcher]. [Further,] the more hurried or careless the process the greater the probability that the bone will [display a butchering mark]” (see also Maltby 1985:22). Gilbert (1979:235) echoed this when he noted that butchering marks were likely created by “the sloppiest efforts at carcass division.” Finally, Maltby (1985) underscores the fact that it is possible that all carcasses represented in a collection were not butchered in like manners. All of these statements presume that although all bones are butchered (speaking metaphorically; see next paragraph), only some of them – for various reasons – in a collection will sustain butchering marks. This presumption as yet has no empirical (actualistic) basis. Nevertheless, when seeking to measure the “intensity” of butchering by quantifying butchering damage, a seldom acknowledged assumption is required.

Most analysts (implicitly) assume that given some set of bones X, some subset X' of those bones will be butchered, and of those butchered bones some subset X'' will sustain damage in the form of butchering marks. The critically important assumption during analysis and interpretation, then, is that some proportion of each category of skeletal part was butchered and some lesser proportion will display butchering marks, and those two proportions will directly and positively covary at least at an ordinal (but likely not a ratio) scale. I am speaking metaphorically when I say that bones are butchered because it is actually carcasses and carcass parts that are butchered, with the notable exception of fracturing of bones for purposes of marrow extraction and grease rendering (e.g., Noe-Nygaard 1977). I use the metaphorical shorthand form *bones are butchered* here for convenience and efficiency.

A fictitious example will make clear the significance of the requisite analytical assumption. Let’s say that there were ten femora and ten humeri available for butchery (X), and all are present in the archaeological collection we are studying. Of those, six femora and five humeri were in fact butchered (X'); say, for example, that flesh was removed from them. Of those butchered elements, for whatever reason(s), only four femora and two humeri display butchery marks (X''). The critical statistical relation here is that more femora than humeri were butchered, and that more femora than humeri display (archaeologically visible) butchery marks. Sixty percent of the observed femora and 50 percent of the observed humeri were butchered, but in fact

only 40 percent of the observed femora and 20 percent of the observed humeri display butchering marks. Thus we could say that femora were more intensively butchered than humeri because proportionally more of the femora than humeri display butchering marks (where *intensity* concerns the amount of energy invested; more butchering marks and more butchery marked bones are thought to signify more energy). But if in fact six of ten femora were butchered and five of ten humeri were butchered, but only one femur displays butchering marks and three humeri display butchering marks, then we would be wrong to conclude that humeri were more intensively butchered than femora (discussion derived from Lyman 1992b, 1995b). Notice that “quantifying butchering damage” was said rather than “quantify butchering marks” or “tally butchery marked bones.” The latter two are often used as synonymous when in fact it should be (and will become) clear that they are quite different.

The preceding discussion focuses on tallying the number of skeletal elements that display butchering marks. This counting procedure mimics those used to quantify burning, corrosion, gnawing, and the like. In all cases, the tallying procedure provides data that answer the question: What proportion (or percentage) of specimens (usually identifiable, or NISP) display a particular kind of taphonomic modification? However, the target variable seems, based on inferences attending observations of percentages of butchery marked specimens, to be the *intensity* of butchering, which is seldom clearly defined but based on published interpretations and a few explicit statements involves the amount of energy spent (e.g., Haynes 2002). Some analysts have therefore worried that tallying the number of butchery marked bones does not actually measure the target variable but something else. These individuals argue that to measure the intensity of butchering, one needs to tally the number of butchering marks so as to have quantitative data that actually reflect the intensity of butchering (Abe et al. 2002; Marean et al. 2001). This is an important observation about the relationship (or lack thereof) between a measured variable (NISP of butchery marked bone) and a target variable (intensity of butchering).

### *Tallying Percussion Damage*

The morphometric criteria for identifying flake scars and percussion damage are spelled out in various places (e.g., Blumenshine and Selvaggio 1988, 1991; Capaldo and Blumenshine 1994; Fisher 1995). Typically, zooarchaeologists have tallied the NISP (NSP more rarely) displaying flake scars, percussion notches, and other percussive damage, and then calculated the proportion or percentage of NISP that displays such marks. Some individuals have tallied the number of marks (e.g., Kooyman

2004), even though it is likely that the number of marks is at least partially a function of the number of specimens examined. Furthermore, some notches overlap one another, and sometimes a single blow will produce a nested series of flake scars (Capaldo and Blumenschine 1994), although perhaps with sufficient training and experience potential difficulties with identifying and counting flake scars and percussion marks might be minimized (Blumenschine et al. 1996). Recent experimental work suggests that tallying percussion damage as the number of distinct marks may be accomplished rather accurately, but the frequency of percussion marks did not correlate with the number of hammerstone blows administered to bone specimens in one set of experiments (Pickering and Egeland 2006). Therefore, for the present, interpreting percussion-mark frequencies (rather than number of specimens with percussion marks) in terms of intensity of butchering (energy invested) is precluded. There is no empirically demonstrable relationship between the target variable and the measured variable.

One might choose to tally the frequency of percussion-damaged specimens across different skeletal elements of a taxon, or across a common set of skeletal elements of several taxa. The analyst might wonder if more humeri specimens, say, display percussion damage than do femora specimens of deer. Alternatively, one might wonder if more long bones of wapiti display percussion damage than do the long bones of deer; wapiti tend to be two to four times larger than deer. In one set of collections, I found that deer long bones had significantly more flake scars than did wapiti long bones, but in another set of collections exactly the opposite situation was found (Lyman 1995b). Why this was the case seemed to relate to the kinds of other resources that were exploited, but lack of actualistic data linking the variables precluded straightforward interpretation.

In sum, there are two basic ways to record percussion damage – as the number of damaged specimens (generally reported as %NISP that has such damage), and the number of instances of force application manifest as individual flake scars, percussion notches, and the like. Explicit statement of a research question will help explicate the target variable and an appropriate measured variable. The relationship between the two, however, may well be unknown, and experimental work is needed in such cases to establish that relationship.

### *Tallying Cut Marks and Cut Marked Specimens*

The morphometric criteria for identifying cut marks are described in numerous places (e.g., Blumenschine et al. 1996; Fisher 1995; Greenfield 1999; Lyman 1987a;

Shipman and Rose 1983), and the identification of such marks is seldom questioned these days. What is receiving the most analytical attention in the first decade of the twenty-first century is how to count cut marks. There are several ways that cut marks have been tallied. Seldom is the proportion of butchery marked skeletal elements (not specimens) determined (see Todd et al. [1997] for an example of tallying cut marked elements). Sometimes, the %NISP that display cut marks is calculated, but that is perhaps not a good procedure given the potential for variation in fragmentation either across the different skeletal elements of a taxon or across different taxa (Abe et al. 2002). What many analysts do is specify some specific anatomical area or portion, whether, say, the distal humerus or the greater trochanter of the femur (an anatomical area or portion) or diaphysis fragment of the tibia, and then determine how many of each of those portions display cut marks (e.g., Guilday et al. 1962; Lyman 1992b; Snyder and Klippel 2003). This assists with keeping track of the anatomical distribution of cut marks. Are they all on diaphyseal pieces, or half on epiphyses and half on diaphyses, and do proportionately more proximal femora have cut marks than distal femora? Thus, if two of five distal humeri display cut marks, then one would conclude that 40 percent of the distal humeri in the collection have butchering marks. The fact that three of those five humeri are complete skeletal elements, one consists of the distal end and distal one third of the diaphysis, and the fifth consists of just the distal condyle is irrelevant to quantifying cut-mark data when they are tallied by an anatomical location of relatively greater or lesser specificity (Lyman 1992b:250).

Other analysts tally the number of individual cut marks. For example, Milo (1998:109) argued that, based on his own butchering experiments, “the relative effort put into cutting in different areas is best reflected by the number of times the tool scored the bone.” But Milo (1998) worried that differential representation of skeletal parts would skew tallies of individual marks. To avoid this sort of problem, Bunn (2001) tallied the total number of cut marks observed on each kind of skeletal element (e.g., humeri, femora). He then divided each kind of skeletal element (he was dealing solely with limb bones) into five, more-or-less equal-sized areas: proximal end, proximal shaft, midshaft, distal shaft, and distal end. The underpinning assumption to the five areas is that cut marks near the ends of long bones likely have something to do with disarticulation whereas those on shafts result from defleshing. Finally, Bunn determined the percentage of all cut marks per kind of skeletal element that occurred in each of the five areas. This is indeed one way to contend with the differential representation of skeletal parts.

There are other ways that the number of cut marks might be tallied and analyzed, but a potentially significant problem attends any such tallying, regardless of how

those tallies are mapped on anatomy or analyzed. If each individual cut mark or single striation is to be tallied, then they must somehow be distinguished for tallying purposes. The problem arises when cut marks overlap. If, for example, a sawing like motion is used – the cutting tool’s edge drags across the bone surface on both push and pull strokes that overlay each other – then strokes producing striae made later in the sequence of strokes may obliterate or at least obscure striae made by earlier strokes. There is no experimental work that evaluates this possibility, and no one has assessed a paleozoologist’s ability to accurately count individual cut marks. But this may be the least of our concerns. For now, however, let us assume that we can tally the number of individual cut marks (each representing a distinct arm stroke). How, then, might we obtain counts of cut marks as opposed to counts of cut marked specimens or tallies of cut marked anatomical areas?

### *The Surface Area Solution*

Recently, the argument has been made that variation in the representation of surface area is relevant to tallying frequencies of cut marks. Abe et al. (2002:650) are concerned about what they refer to as the “fragmentation dilemma.” In particular, they are worried that “fragmentation generally decreases the number of cut marked fragments and cut mark counts relative to total fragments” (Abe et al. 2002:649). Fragmentation generally means breakage such that what was a single discrete object after fragmentation comprises multiple discrete objects (Lyman 1994c:509). *Fragmentation* destroys the original integrity of a discrete object, but the material or substance comprising that discrete object still exists and, importantly, some of the original integrity of the discrete object may also remain. *Destruction* generally means the complete loss of the original integrity of the object, such as when the specimen is crushed into dust or into pieces that are so small that they cannot be identified and thus are analytically invisible. Abe et al. (2002:649) state “the fragmentation process moves fragments into the unidentifiable category and destroys less-dense bone altogether.” They have in mind extreme fragmentation or destruction in the sense that the specimen is analytically invisible. This process was recognized long ago by Watson (1972; see also Lyman and O’Brien 1987). Less extreme fragmentation, such as when a bone specimen is broken into two or three pieces that can be identified to skeletal portion, may increase the number of cut marked specimens if a fracture plane truncates a cut mark such that half of the mark occurs on one specimen and the other half occurs on another specimen. Refitting specimens is the only way to correct for this.

Given their concern about the destruction of cut marks, Abe et al. (2002:650) suggest “the likelihood of a cut mark being preserved and counted by an analyst is a function of the amount of bone surface area studied and recorded.” This is commonsensical – the more surface area examined, the more cut marks will be found. Abe et al. (2002:650) use this observation, however, to argue that (i) because more cut marks will be found if more surface area is examined, (ii) if we determine the density of cut marks per unit of surface area in an anatomical region, (iii) then “we can correct the number of cut marks by the amount of examined surface area, much as demographers standardize population size by estimating population density.” In particular, they assume that if half (50 percent) of the potential surface area of an anatomical region has been examined, and ten cut marks have been tallied, then were 100 percent of the surface area of that region examined, twenty cut marks would be tallied. They are assuming that the density of cut marks on an observed sample is the density of cut marks on the unobserved remainder of the population. They are assuming precisely what they are trying to discover – the original (predestruction) frequency of cut marks (see also Lyman 2005b).

Abe et al. (2002) are trying to take advantage of the visibility of one variable – frequencies of cut marks on observable bone surface area – in order to measure a variable that is invisible – cut mark frequencies on missing or destroyed bone surfaces. There are a plethora of problems with this procedure. The first problem is that one must decide what comprises the sample of specimens to be examined for any given analysis, and thus one must decide how to define an aggregate of remains. If different specimens are included in a sample, different results are likely to attend analysis of cut mark frequencies. The second problem concerns how to define anatomical regions for which the amount of observed surface area will be determined. The proportion of observed surface area is based on the maximum MNE for any given anatomical region (e.g., proximal end, proximal shaft, mid shaft). Thus, if there is evidence of ten distal humeri (= MNEmax), then there should be ten proximal humeri represented even if only two are observed. But, are proximal ends and distal ends, proximal shafts and distal shafts and mid shafts, such as proposed by Abe et al. (2002), appropriate tallying units? That is presently unclear.

The third problem is that the analytical procedure ignores the historically contingent nature of butchering episodes (Lyman 1987a, 2005b). It is easy to show that even in experimentally controlled situations, for reasons that are unclear, there is a tremendous range of variation in the frequency of cut marks generated in any given butchering episode. Consider the experimental data generated by Pobiner and Braun (2005) and summarized in Table 7.4. Those data are the number of cut marks generated during the defleshing of six goat hindlimbs (femora and tibiae). Each hindlimb

Table 7.4. Frequencies of cut marks per anatomical area on six experimentally butchered goat (*Capra hircus*) hindlimbs. %FR, amount of flesh removed from femur prior to butchering. N-CM, number of cut marks; P, proximal end; PS, proximal shaft; MS, mid shaft; DS, distal shaft; D, distal end. Data from Pobiner and Braun (2005)

Limb	Element	% FR	N-CM P	N-CM PS	N-CM MS	N-CM DS	N-CM D
1	Femur	50	0	0	13	15	0
2	Femur	50	0	3	11	0	0
3	Femur	25	0	1	8	3	0
4	Femur	25	0	0	0	5	3
5	Femur	0	0	5	21	2	0
6	Femur	0	22	6	6	19	0
1	Tibia	0	0	0	0	0	0
2	Tibia	0	0	2	0	5	0
3	Tibia	0	0	2	0	0	0
4	Tibia	0	0	7	13	0	0
5	Tibia	0	0	20	3	0	0
6	Tibia	0	0	0	0	10	0

was defleshed independently of every other hindlimb, and although the amount of flesh on the femur varied when each butchery event began, nothing else did. If Abe et al. (2002) are correct that the observed density of cut marks (frequency per unit area) can be used to estimate the frequency of cut marks that have been destroyed, then there should be minimal variation in the number of cut marks per anatomical region described in Table 7.4, given that those anatomical regions are identical from specimen to specimen in terms of surface area.

The data in Table 7.4 indicate that there is a great deal of variation in the density of cut marks, or the number of cut marks per unit of surface area even when long bones are treated as comprising five distinct regions (proximal and distal ends, proximal and distal shafts, mid shaft). And this is so regardless of whether the amount of meat on a bone was similar from case to case or was different from case to case. Frequencies of cut marks in a given region on individual femurs range from zero to twenty-two (proximal femur), and on individual regions of tibiae they range from zero to twenty (proximal shaft). Given that the amount of surface area of, say, the proximal tibia shaft does not vary significantly across the six specimens, following Abe et al.'s (2002) suggested procedure, were only the proximal shaft of tibia five

recovered, its twenty cut marks would suggest that there were twenty cut marks on each of the other missing proximal shafts of tibia (based on an MNE of six total recovered distal tibiae). Data in Table 7.4 indicate that such an inference is incorrect.

The fourth problem that attends determination of the number of missing cut marks based on observable frequencies of cut marks per unit of surface area is that it is not at all clear what the visible frequencies of cut marks are measuring. Abe et al. (2002:657) state that a “key assumption that all zooarchaeologists make is that more intensive cutting (more cutting actions) results in higher frequencies of cutmarks on the bone surface.” This is indeed a key assumption. Given that creating two cut marks requires two arm strokes, but creating one cut mark requires one arm stroke, it is likely that what most analysts mean by intensity is number of arm strokes. The analytical assumption in a paleozoological context, then, must be that as the number of arm strokes or slices increases, so too does the number of cut marks created. Unfortunately, experiments by Egeland (2003) indicate that there is no relationship between the number of arm strokes used to butcher limbs of large mammals and the number of cut marks that are generated.

Egeland (2003) butchered sixteen partial and complete limbs (fore and hind) of domestic cows (*Bos taurus*) and domestic horses (*Equus caballus*). Stone tools were used to remove flesh, arm strokes aimed at flesh removal were tallied, and the amount of flesh removed was recorded (Table 7.5). There is neither a statistically significant relationship between the number of arm strokes and the number of cut marks created across the ten multiskeletal element limbs Egeland butchered ( $r = -0.206, p = 0.52$ ), nor is there a statistically significant relationship between the number of arm strokes and the number of cut marks created across the 31 individual skeletal elements Egeland butchered ( $r = -0.20, p = 0.28$ ). These results do not change if the data are log-transformed (Figure 7.5). This means that when we tally cut marks, we cannot conclude that more cut marks on skeletal parts comprising the ankle joint than on skeletal parts comprising the wrist joint means that the ankle was more intensively butchered than the wrist. There is no actualistic research indicating the validity of the relationship between the two variables (measured = number of cut marks; target = number of arm strokes or intensity) and there are actualistic data (Egeland’s) which show that at least some times there is no such relationship at all.

In sum, then, the surface area solution proposed by Abe et al. (2002), although perhaps solving various problems that attend tallying the number of specimens that have cut marks, introduces problems of its own. It is dependent on the aggregate of specimens included, it is dependent on how skeletal regions are defined, it ignores the historically contingent and variable process of butchering, and one ultimately assumes what one is trying to ascertain. The last is so because the analytical protocol

Table 7.5. *Frequencies of arm strokes and cut marks on sixteen limbs of cows and horses. Number in first column identifies the unique butchering episode. Data from Egeland (2003)*

Limb/Element	Taxon	N of Strokes	N of Cut Marks	Meat Removed (kg)
1 hindlimb	cow	3747	11	—
2 forelimb/scapula	horse	535	8	8.60
2 forelimb/humerus	horse	877	7	7.10
2 forelimb/radius-ulna	horse	525	44	2.80
3 hindlimb/tibia	horse	577	8	3.40
4 hindlimb/tibia	horse	582	14	2.50
5 hindlimb/tibia	horse	594	1	3.80
6 hindlimb/tibia	horse	202	1	0.50
7 hindlimb/femur	horse	2155	3	25.7
7 hindlimb/tibia	horse	420	29	4.5
8 hindlimb/femur	horse	1757	0	17.5
8 hindlimb/tibia	horse	650	2	2.9
11 hindlimb/femur	cow	687	31	17.4
11 hindlimb/tibia	cow	715	22	4.1
12 forelimb/scapula	cow	395	5	6.1
12 forelimb/humerus	cow	371	7	4.8
12 forelimb/radius-ulna	cow	362	8	1.9
13 forelimb/scapula	horse	739	26	3.1
13 forelimb/humerus	horse	1124	9	4.4
13 forelimb/radius-ulna	horse	586	4	2.1
14 forelimb/scapula	horse	5397	13	8.4
14 forelimb/humerus	horse	2265	17	5.5
14 forelimb/radius-ulna	horse	2080	9	2.4
15 forelimb/scapula	cow	986	0	3.0
15 forelimb/humerus	cow	532	0	6.3
15 forelimb/radius-ulna	cow	951	0	2.3
19 forelimb/scapula	cow	148	20	1.1
19 forelimb/radius-ulna	cow	178	33	0.7
21 forelimb/scapula	cow	596	31	5.2
21 forelimb/radius-ulna	cow	695	31	2.4
22 forelimb/scapula	cow	502	107	5.8
22 forelimb/radius-ulna	cow	877	29	2.3

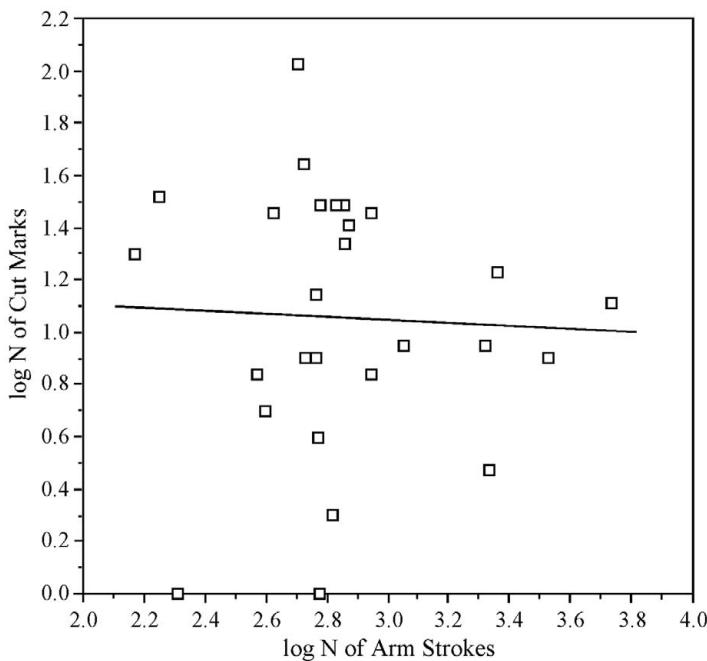


FIGURE 7.5. Relationship between number of arm strokes and number of cut marks on thirty-one skeletal elements ( $r = -0.235$ ,  $p = 0.2$ ). Data from Table 7.5.

demands the assumption that a sample of bone surface area gives an accurate estimate of the density of cut marks across the total (population's) surface area of bones, whether those bones are present, destroyed, or not collected. This might be so if cut marks were randomly distributed across bone surfaces, but this is unlikely to be true for many reasons and empirical data indicate it is not true. In short, we cannot assume what we are trying to discover.

## DISCUSSION

One driving force behind study of the frequencies of bone specimens displaying butchery damage and frequencies of specimens displaying carnivore damage concerns the roles of meat eating and of carcass acquisition (hunting or scavenging) in hominid evolution (see Domínguez-Rodrigo [2002] and Lupo and O'Connell [2002] for recent reviews). The underlying assumption comprises two interrelated parts. First, if carnivores have access to a prey carcass before tool-carrying butchers, prey bones will have many tooth marks but few butchering marks; if hominid butchers have access to a prey carcass prior to access by a carnivore/scavenger, then

bones of prey will have many butchering marks (especially cut marks representing defleshing of meat-rich proximal limb elements) and few tooth marks of carnivores. Second, the more flesh on bones, the more cut marks are expected. Each of the preceding statements is carefully phrased; each refers to the frequency of marks, not the frequency of marked skeletal elements or skeletal specimens. The target variable is clear – how many marks are there per specimen – the significant assumption is that more flesh results in more marks, whether cut marks or tooth marks. Variation in the relative frequencies of the two kinds of marks depends on order of access and the amount of flesh remaining that the second carnivore (whether a quadruped or biped) can exploit.

The critical assumption is worded so as to emphasize that frequencies of marks – whether butchering marks or tooth (gnawing) marks – is the critical variable, that is in fact also how individuals who have debated the issue phrase the assumption (e.g., Binford 1986, 1988; Bunn and Kroll 1986, 1988; Domínguez-Rodrigo 2002; Lupo and O’Connell 2002; Pobiner and Braun 2005; Selvaggio 1994, 1998; Thompson 2005). But almost without fail, paleozoologists involved in the discussion do not tally up cut marks and tooth marks; instead they tally up cut marked bones and tooth marked bones and analyze those frequencies. The relationship between the number of marked bones (measured variable) and the property or process of interest (target variable) is obscure. Furthermore, the target variable is inexplicit – is it the amount of meat associated with a bone, the size of the carcass, the size of the bone – and this contributes to the obscure relationship between it and the measured variable. Some examples will make this clear.

Among the data in Table 7.5, the number of strokes necessary to deflesh an individual skeletal element is significantly correlated with the amount of meat removed from the element ( $r = 0.365, p = 0.044$ ), especially if both variables are log transformed ( $r = 0.592, p = 0.0005$ ; Figure 7.6). The number of cut marks per skeletal part, however, is not correlated with the amount of meat removed from a bone ( $r = -0.079, p = 0.67$ ), and this holds true for the log transformed data as well (Figure 7.7). These results suggest the interpretive assumption that more cut marks means there was more meat on bones for stone-tool wielding butchers to remove is unfounded. However, the data in Table 7.6 may support the assumption. Those data were generated by Pobiner and Braun (2005), who provided eighteen hindlimbs comprising the femur and tibia to stone-tool wielding butchers. But, before the limbs were turned over to the butchers, different amounts of flesh were removed by Pobiner and Braun to simulate early access to fully fleshed limbs, later access to partially fleshed limbs (25 percent of flesh removed prior to butchery), and still later access to rather defleshed limbs (50 percent of flesh removed prior to butchery).

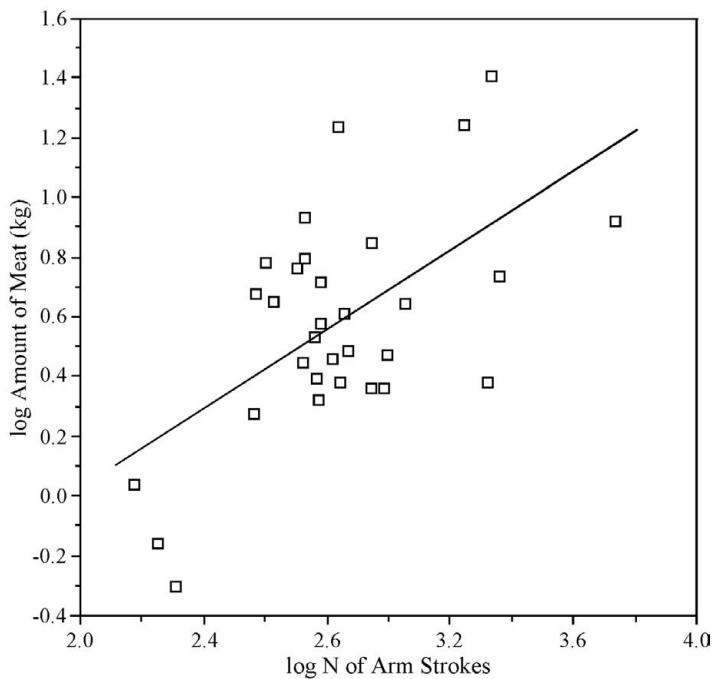


FIGURE 7.6. Relationship between number of arm strokes necessary to deflesh a bone and the amount of flesh removed ( $r = 0.592$ ,  $p = 0.0005$ ). Data from Table 7.5.

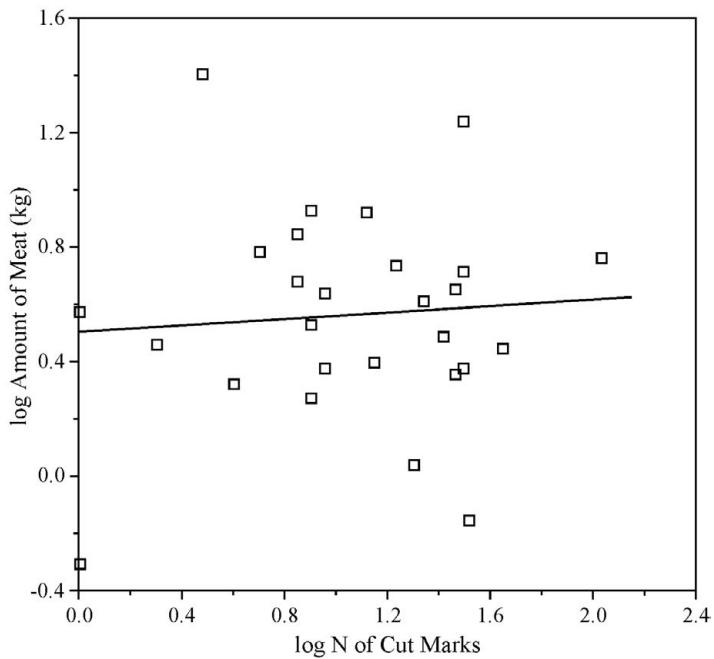


FIGURE 7.7. Relationship between number of cut marks and the amount of flesh removed from thirty-one limb bones ( $r = -0.035$ ,  $p = 0.85$ ). Data from Table 7.5.

Table 7.6. Number of cut marks generated and amount of meat removed from eighteen mammal hindlimbs (femur + tibia) by butchering. Data from Pobiner and Braun (2005)

Limb	Taxon	Meat removed (kg)*	N of cut marks
1	cow (juvenile)	14.25	15
2	cow (juvenile)	12.00	14
3	cow (juvenile)	4.25	56
4	cow (juvenile)	3.25	15
5	cow (juvenile)	1.00	20
6	cow (juvenile)	0.50	69
7	goat	0.390	31
8	goat	0.538	14
9	goat	0.814	12
10	goat	0.786	8
11	goat	1.084	28
12	goat	1.010	53
13	zebra	23.0	73
14	zebra	23.5	67
15	zebra	13.5	90
16	zebra	14.0	160
17	zebra	23.0	95
18	zebra	23.0	45

\* Note, the amount of meat removed varies intrataxonomically because of prebutchery meat removal aimed at testing the hypothesis that the amount of meat remaining for removal would correlate with the number of cut marks.

Pobiner and Braun (2005) conclude that there is no relationship between the number of cut marks and amount of meat removed *within each size class of butchered animal*. Their data on amount of meat removed and number of cut marks produced per each of eighteen butchering experiments are plotted in Figure 7.8 by taxon. There is no significant relationship between the two variables intrataxonically for any of the three taxa represented (goat,  $r = 0.62$ ,  $p = 0.19$ ; cow,  $r = -0.78$ ,  $p = 0.065$ ; zebra,  $r = -0.48$ ,  $p = 0.34$ ). However, there is a positive relationship between carcass size and number of cut marks when all carcasses were included and analysis is intertaxonomic rather than intrataxonomic (Figure 7.9,  $r = 0.49$ ,  $p = 0.039$ ). It matters little which analysis is correct. The varied results highlight a critical point. Target variables must

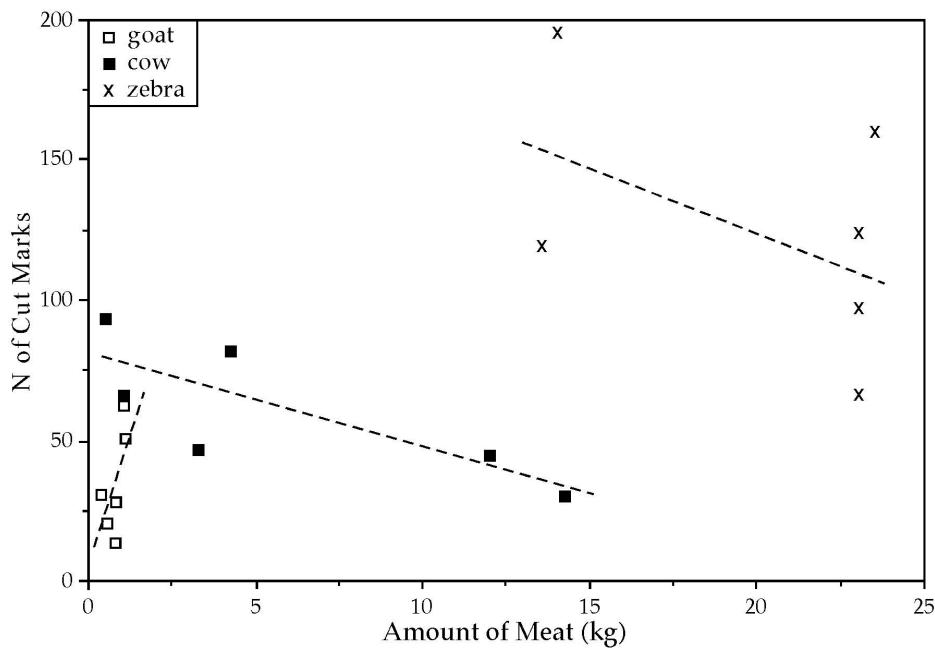


FIGURE 7.8. Relationships between number of cut marks and the amount of flesh removed from six hindlimbs in each of three carcass sizes. Simple best-fit regression lines (dashed) are shown for each carcass size. None of the relationships is statistically significant (goat,  $r = 0.62, p = 0.19$ ; cow,  $r = -0.78, p = 0.065$ ; zebra,  $r = -0.48, p = 0.34$ ). Data from Table 7.6.

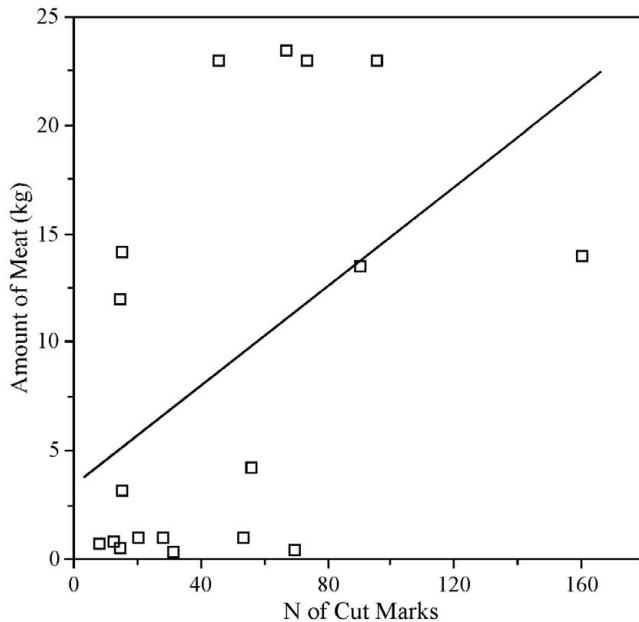


FIGURE 7.9. Relationship between number of cut marks and the amount of flesh removed from eighteen hindlimbs ( $r = 0.49, p = 0.039$ ). Data from Table 7.6.

be explicitly defined, as must measured variables and the suspected relationship between the two. Such specifications will assist with determination of the appropriate statistical tests.

With respect to quantifying cut marks, obscure target variables and poorly understood relationships between target and measured variables are not the only aspects of the quantitative data that are recorded, analyzed, and reported that likely contribute to the lack of resolution of the debate whether early hominids hunted large game or merely scavenged long-dead carcasses. Lupo and O'Connell (2002:102) correctly note that analysts report tallies of cut marked specimens (and tooth marked specimens) differently. Many analysts report the number of marked specimens per portion (e.g., proximal, distal, shaft) per skeletal element (e.g., humerus, radius, femur) (e.g., Domínguez-Rodrigo 1997; Domínguez-Rodrigo and Pickering 2003); a few report the number of marked specimens per skeletal element with no distinction of portion of element (e.g., Domínguez-Rodrigo 1999a; Oliver 1994); and a few report the number of marked specimens per portion (proximal, shaft, distal) of skeletal element with no distinction of skeletal element (e.g., Blumenschine 1995; Capaldo 1997). It is these sorts of ambiguities that in part prompted a flurry of rebuttals and responses regarding interpretations of the cut mark and tooth mark data (e.g., Domínguez-Rodrigo 1999b, 2003a, 2003b; Monahan 1999; O'Connell and Lupo 2003; O'Connell et al. 2003). A major cause of the debate has been poorly developed and weakly warranted methods that are incompletely described – the problem identified by Domínguez-Rodrigo – in conjunction with poorly worded and incompletely developed theoretically informed interpretive models – the problem identified by O'Connell et al. (2003). In terms used throughout this volume, measured variables are inexplicit and have at best a poorly understood relationship to target variables.

## CONCLUSION

Discussions in other arenas summarize in somewhat different terms what has been discussed in this chapter. With respect to attributes on prey remains created by predators, Kowalewski (2002:14) states that “the frequency of traces is arguably the most important and widely used metric in quantitative analyses of the fossil record of predation that estimates the frequency of predator–prey interactions and may serve as a proxy for predation intensity.” But when he describes ways to tally the frequency of traces, he in fact suggests the number of specimens with traces be tallied and thus correctly notes that “the number of specimens with traces of predation is *not* synonymous with the total number of traces found in those specimens unless all specimens

bear singular traces. When computing predation intensity we should always use the number of prey specimens attacked (i.e., the number of specimens with traces) and not the number of attacks (i.e., the number of traces)" (Kowalewski 2002:15). This is because the target variable is predation intensity, implied by Kowalewski to comprise the fraction of the prey population that has in fact been preyed on. Tallying numbers of predation marks would thus not measure the frequency or intensity of predation but rather how often a particular prey organism was attacked.

Kowalewski (2002) is concerned with organisms that have single element skeletons, and so he notes that measuring predation intensity may require modification to measurement techniques if skeletons of prey comprise multiple elements. This is so for the same reason that the number of traces would not measure the intensity of predation but rather the frequency of attacks (individual prey may be attacked more than once). Counting predation traces on multiple but different skeletal elements introduces the problem of interdependence – has one attack been counted more than once because multiple elements of a single organism have been tallied? Measuring the intensity of carnivore gnawing, corrosion, burning, and butchering, however, because of how "intensity" is (typically implicitly) defined, requires tally of those potentially interdependent specimens.

Among other approaches to mapping predation traces on the anatomy of the prey skeleton, Kowalewski (2002:25) distinguishes a "qualitative approach" and a "sector approach." The first involves mapping each trace on a single standard skeletal element. Although this approach precludes statistical comparison of data sets and is subject to mapping error based on operator error and morphological and allometric variation among specimens, it does reveal anatomical areas that may have taphonomic or biological significance. It has been used by various taphonomists. The sector approach involves partitioning the skeleton or skeletal elements into sectors and tallying the number of traces in each. This approach allows statistical comparison of data sets, such as  $\chi^2$  analysis and calculation of evenness and heterogeneity indices. This approach too has been used by various taphonomists. At the risk of being redundant, the approach chosen should be dictated by the research question.

Discussion in this chapter is not to resolve debates over whether early hominids were scavengers, hunters, or acquired meat using both techniques. Rather, the goals of the chapter have been two. First, methods used to quantify various sorts of damage to bones – weathering, corrosion, gnawing, burning, butchering – have been described. Second, the critically significant nature of the relationship between a measured variable and a target variable and the critically significant fact that each variable must be explicitly defined have been highlighted. Precisely the same (then ambiguous) relationship underpinned debates in the 1950s through 1980s regarding the relationship

between NISP, MNI, biomass, and other measures of taxonomic abundances, and a target variable of abundances of taxa exploited by people or abundances of taxa on the landscape (Chapters 2 and 3). Those debates were more or less resolved in the 1980s as two things became clear. First, any measure of taxonomic abundance was found to be at best ordinal scale (or to be an estimate), and second, the relationship between a chosen measured variable (NISP, MNI, biomass) and the target variable was a taphonomic question. Many paleobiologists came to both conclusions using “fidelity studies,” actualistic research on the relationship between recently formed assemblages of faunal remains and the accuracy with which they reflect taxonomic abundances in the faunas from which the collections derive (see Chapter 2 for a formal definition of fidelity studies). The success of these studies resides in unambiguous definitions of measured and target variables. Ambiguity with respect to measured variables and target variables permeates many modern taphonomic studies. It is no wonder that we do not understand the relationship between two variables when one or more of them is poorly defined or is simply inexplicit.

## Final Thoughts

In this volume, some of the most basic issues of quantifying different kinds of paleozoological data have been explored. A bit more than two decades ago, Grayson (1984) published a book-length treatment on the same general topic, and that seemed to resolve many of the debates over how to quantify taxonomic abundances. Arguments over whether NISP or MNI was the better measure nearly ceased to appear in the literature. Yet, some individuals continue to report MNI values, either as the unit of choice for quantifying taxonomic abundances (e.g., Avery 1991, 1992; Landon 1996), or apparently for the sake of complete descriptive reporting (e.g., Plug 2004; Stahl and Athens 2001). A few continue to develop innovative ways of tallying MNI (e.g., Vasileiadou et al. 2007). The usual reason given for use of MNI is that NISP is subject to intertaxonomic variation in fragmentation and so gives potentially biased estimates of taxonomic abundances. Although it is true that NISP *can* influence estimates of taxonomic abundance, those who use the differential fragmentation argument as a warrant to determine MNI values neither empirically evaluate the truthfulness of this warrant in their particular instances nor fail to present NISP data. Why do they present what they take to be biased data? Why do they not determine if in fact fragmentation varies intertaxonomically rather than simply assert that it does? Perhaps they do not because of a lack of mathematical and statistical sophistication. That lack of sophistication is a major reason for this book.

Some have argued on the basis of ethnoarchaeological (Hudson 1990) or historical (Breitburg 1991) data that MNI provides more accurate estimates of taxonomic abundances than NISP. That may well be so in particular cases where aggregation and derivation of MNI is not dependent on analytical choices; we must make these choices when dealing with prehistoric materials. Reitz and Wing (1999:199) state that MNI is the “only way to compare mammals, birds, reptiles, amphibians, fishes, and mollusks,” but the arguments in Chapter 2 identify the fallacious nature of this statement. Given the continued use and advocacy of MNI, arguments made

by Richard Casteel and Donald Grayson regarding the nature of MNI and its statistical relationship with NISP, as well as their characterizations of MNI and NISP as quantitative units, have been reiterated for a new generation of paleozoologists. This is not to say that MNI is always the wrong quantitative unit to use. Both logic and empirical data indicate, however, that it typically is the wrong unit to use when some measure of taxonomic abundances is needed. It has been argued that MNE is not as good a quantitative unit as NISP when one needs a measure of skeletal part abundances.

Other methods of quantifying taxonomic abundances, such as estimating biomass, have also been reviewed. Some analysts continue to calculate meat weight using Theodore White's method (references in Dean 2005b). (Ornithologists still use the Whitean method of multiplying the MNI of prey evident in a sample of egested pellets by the average weight of an individual prey to determine biomass [Leonardi and Dell'Arte 2006].) The skeletal mass allometry technique is quite popular in some areas, and it continues to be used today (e.g., Carder et al. 2004; Lapham 2005; Pavao-Zuckerman 2007). Chapter 3 of this volume was written with the express purpose of highlighting some of the weaknesses of estimating biomass. Because many of the quantitative variables paleozoologists seek to measure are dependent on sample size, Chapter 4 summarizes the various ways that sample-size effects might be detected and analytically controlled. Chapter 5 covers a central issue in paleozoology—quantifying and comparing the structure and composition of prehistoric faunas, and monitoring trends in taxonomic abundances. Chapter 6 provides detailed coverage of a quantitative unit that has been extensively used over the past 20+ years – MNE – even though it has been around virtually as long as MNI. And Chapter 7 describes ways to tally and analyze quantitative variables that concern taphonomic agents and processes. What could possibly be left to discuss?

There is one thing that warrants comment. This concerns the fact that statisticians have found it necessary to comment on quantitative paleozoology. This commentary began with Ringrose's (1993) detailed discussion that is still quite worthwhile to read. Pilgram and Marshall (1995) pointed out that Ringrose apparently had little experience with faunal remains, and so some of his comments were a bit off base. Ringrose (1995) responded that although he did not in fact know very much about the realities of paleozoology, he commented in kind that Pilgram and Marshall (1995; Marshall and Pilgram 1991) seemed to not be as statistically sophisticated as he (at least) hoped paleozoologists might be. It was with that discussion firmly in mind that I have included minimal discussion of statistics and focused on what simple statistical analyses might reveal about the quantitative properties of a paleozoological

collection. In an effort to make revelations clear, graphs of statistical relationships are included, along with various statistical results attending the graphed relationships. And, in most cases, the data underpinning the graphs and the statistics are included to allow the interested reader to replicate analyses graphically and statistically. Replication will assist comprehension of an analytical technique, and it ensures correct implementation of the technique. Hopefully, readers will find utility in the many graphs and tables.

Paleozoologists who read this volume may well hope for more, or less, statistical sophistication. Not being a statistician, I can only reply: Read a statistics book. But in saying that, I also want to make the observation that, like Ringrose (1993), other statisticians have contributed to the discussions on quantitative paleozoology. And it is clear that at least some of those statisticians are, like Ringrose, not aware of the practical realities of quantitative paleozoology. Thus, MNE is (incorrectly) defined as “the NISP calculated for each skeletal part” (Baxter 2003:212) by a statistician. Such errors are not restricted to those who are not paleozoologists. NISP has been said by paleozoologists to be the Number of Identified Skeletal Parts, or the Number of Identified Skeletal Portions, yet they do not define *skeletal part* or *skeletal portion*. Such loose use of key terms is commonplace in many scientific endeavors, but that does not make it acceptable. Explicitly defined terminology is critical to the success of any research; such is all the more critical with respect to quantitative units, whether fundamental or derived. That NISP has various definitions (or at least descriptions) in the literature reflects poor understanding of the term “specimen” and how it compares with “skeletal element” and the generic “bone.” Using the definitions in Chapter 1 of this book, or a similar set of definitions that are explicitly stated by the researcher, would help the discipline a lot.

This is not a book about terminology. It is instead a book about how to count faunal remains – bones, teeth, shells, and fragments thereof. To reiterate, one should read a statistics book to learn about statistics; read *Quantitative Paleozoology* to learn about counting faunal remains. In so doing, and putting the two together, a paleozoologist may well conceive of a unique analysis that reveals something about the behavior of a quantitative unit or gain insights to some aspect of a collection of broken bones. In most chapters, knowledge about the relationship (or lack thereof) between a target variable and a measured variable has been emphasized. In many cases, such knowledge is crucial to valid interpretation, but it may not always be required. In some cases, exploratory data analysis may suggest further analyses are necessary because of a particular relationship between two variables. The nature of the relationship between variables may suggest other sorts of variables that need to

be measured in order to understand the relationship. As a way to conclude this book, I outline an example.

#### COUNTING AS EXPLORATION

Grayson (1979, 1984) suggested that analyses of the relationship between NISP and MNI might prove revealing. Such analyses might reveal something about the particular collections studied, something about the nature of the relationship between these two most basic counting units, or both. Recall that Klein and Cruz-Uribe (1984) noted that their results were different than Casteel's (1977, n.d.) with respect to the statistical relationship they found between NISP and MNI. Because of that difference, Klein and Cruz-Uribe suggested that perhaps the set of assemblages they had used to examine the relationship comprised remains that were much more fragmented than those remains in the assemblages that Casteel had used. This was an astute observation to make, but it was also one that Klein and Cruz-Uribe could not evaluate empirically given a lack of appropriate data. They did not have NISP:MNE data for the various skeletal elements because the data they used (and those used by Casteel) were derived from literature that did not present that data (it was not a target variable of the analysts). About the same time that Klein and Cruz-Uribe (1984) presented their conclusion, Bobrowsky (1982) pointed out that Casteel (1977, n.d.) had lumped numerous taxa together, and that such lumping masked the influence of intertaxonomic variation in the number of identifiable elements per individual skeleton. That is, Bobrowsky identified a cause for variation in the relationship between NISP–MNI data pairs that was different than the cause identified by Klein and Cruz-Uribe.

In a clever bit of analysis, Bobrowsky (1982) chose one stratum from one site and compared the relationship of NISP to MNI across four taxonomic groups (birds, mammals, reptiles, fish) represented by the remains from that single stratum. His results indicate that indeed, intertaxonomic variation in the number of identifiable elements per individual skeleton significantly influenced the slope of the best-fit regression line described by the model in Figure 2.4. Thus, the line describing the relationship between the NISP and MNI data from remains of birds had a steeper slope and higher plateau (it leveled at a higher MNI) than did the line for mammals. Bobrowsky (1982) found this relationship between the two lines expectable given that each bird skeleton tended to provide fewer taxonomically identifiable elements than did each mammal skeleton. This simply meant that each additional NISP of birds was more likely to contribute a new MNI than was each additional NISP of mammals. I can identify to the genus or species level about forty-five to forty-eight kinds of

Table 8.1. Statistical summary of relationship between NISP and MNI in collections of paleontological birds, paleontological mammals, archaeological birds, and archaeological mammals.  $p < 0.0001$  in all

	N of assemblages	N of data pairs	Pearson's $r$	$r^2$	Slope	Y intercept
Paleontological birds	7	265	0.8747	0.7651	0.483	-0.05036
Paleontological mammals	11	360	0.8719	0.7586	0.5581	-0.09384
Archaeological birds	22	696	0.9133	0.8342	0.629	-0.07764
Archaeological mammals	35	764	0.8963	0.8034	0.5561	-0.06826

skeletal elements (including isolated teeth, ignoring side differences and fragments) of a typical mammal skeleton. In a detailed study of an archaeological avifauna, Broughton (2004) identified sixteen kinds of skeletal elements across forty-six genera. This anecdotal information suggests Bobrowsky (1982) may have been correct.

Klein and Cruz-Uribe's (1984) concern about fragmentation, and Bobrowsky's (1982) concern about intertaxonomic variation in the number of identifiable elements per skeleton both concern variables that influence the ratio NISP:MNI. Keeping these variables in mind, I compiled NISP–MNI data pairs for paleozoological assemblages in North America (the data to which I have the easiest access). To keep the intertaxonomic variable simple, I compiled data for only birds and mammals. I also compiled and kept separate data for both paleontological and archaeological avian and mammalian assemblages. My reason for doing so was that it seemed reasonable to suppose that remains of animals from archaeological assemblages, particularly those of mammals but perhaps not those of birds, would be more fragmented than the faunal remains in paleontological collections. It is, after all, well known that human butchers tend to break bones with some regularity (e.g., Noe-Nygaard 1977; Thomas 1971). By definition a human taphonomic agent had not influenced paleontological collections.

Descriptive statistics for the four sets of data are summarized in Table 8.1. There are several things that need to be considered here. First, graphs of the relationships between NISP and MNI in the four assemblages indicate that the log-transformed data describe a straight line. The straight-line relationship is apparent for the paleontological bird remains (Figure 8.1), the paleontological mammal remains (Figure 8.2),

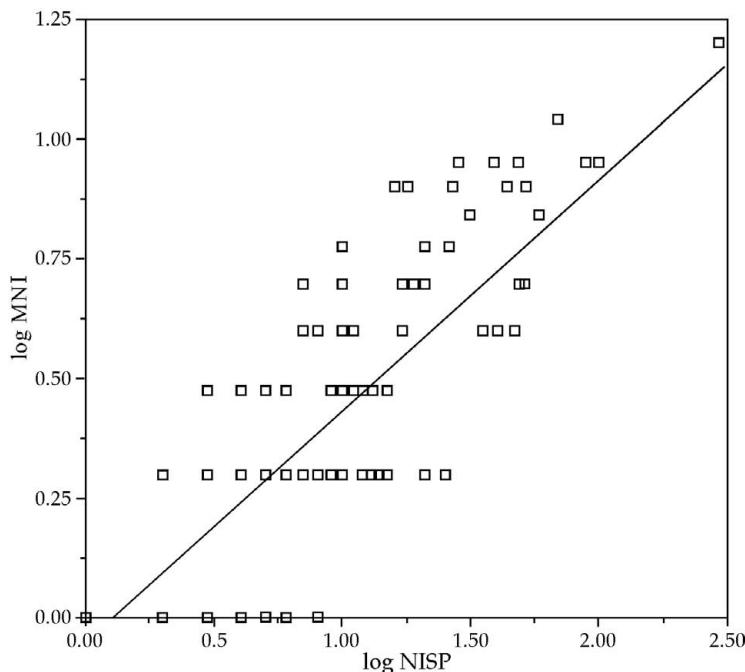


FIGURE 8.1. Relationship between NISP and MNI in seven paleontological assemblages of bird remains from North America. The number of data points is 265; not all are visible because of duplication and overlap.

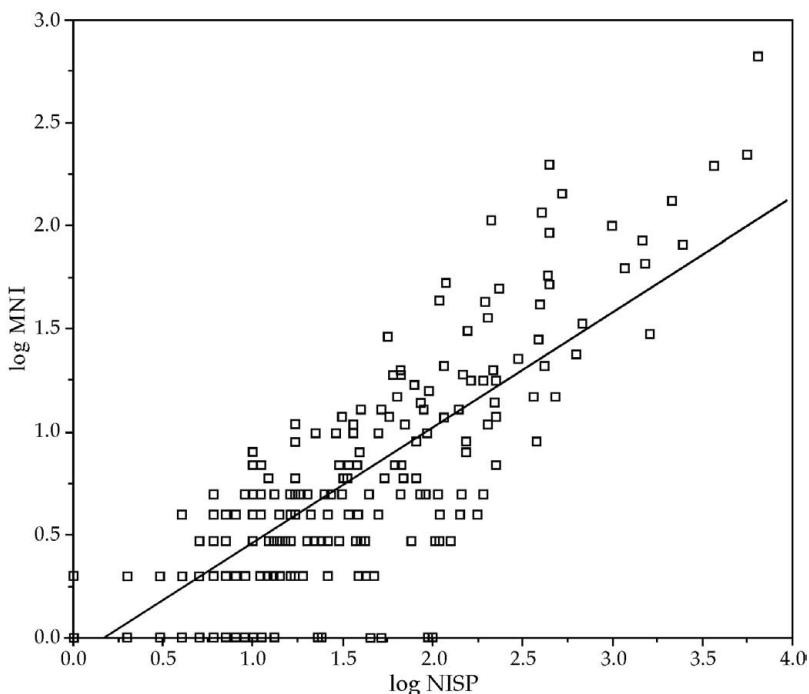


FIGURE 8.2. Relationship between NISP and MNI in eleven paleontological assemblages of mammal remains from North America. The number of data points is 360; not all are visible because of duplication and overlap.

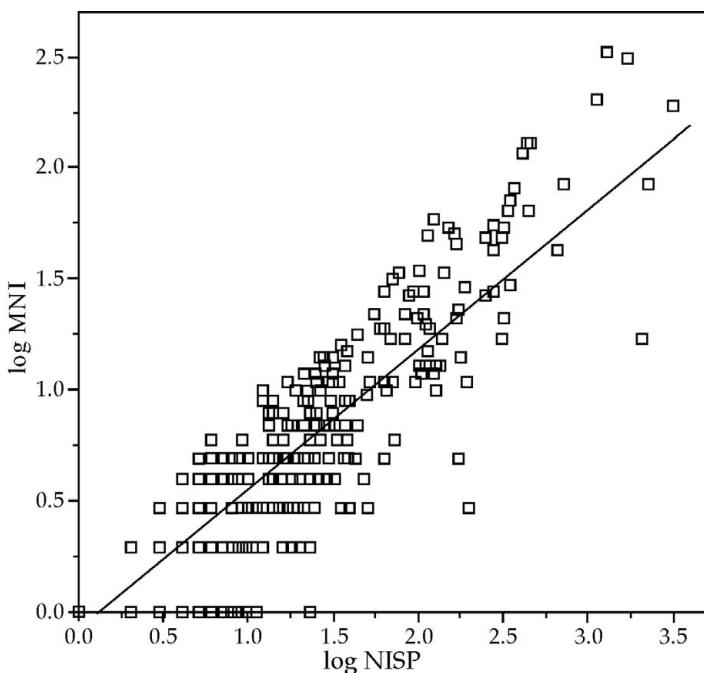


FIGURE 8.3. Relationship between NISP and MNI in twenty-two archaeological assemblages of bird remains from North America. The number of data points is 696; not all are visible because of duplication and overlap.

the archaeological avian remains (Figure 8.3), and the archaeological mammal remains (Figure 8.4). The lines plotted in these graphs should by now be familiar; they are simple best-fit regression lines described by the formula  $Y = aX^b$ , where  $X$  is the independent variable ( $\log \text{NISP}$ ),  $Y$  is the dependent variable ( $\log \text{MNI}$ ),  $a$  is the  $Y$  intercept (it should be zero, given that a zero value for NISP must produce a zero value for MNI; note that all empirically determined values are quite close to zero [Table 8.1]), and  $b$  is the slope of the line. Variables  $a$  and  $b$  are constants determined empirically for each data set. In all cases, the relationship between NISP and MNI is statistically significant ( $p < 0.0001$ ) and variation in NISP explains 76 percent or more ( $= r^2$ ) of the variation in MNI.

The data in Figures 8.1–8.4 mimic the results of the data used by Bobrowsky, Casteel, Grayson, and Hesse; a tight statistical relationship between NISP and MNI is apparent. Clearly, MNI would seem to always increase as NISP increases. The slope of the line (Table 8.1) (which measures the rate of change in MNI relative to the rate of change in NISP) for paleontological mammals ( $b = 0.5581$ ) is not significantly different from that for archaeological mammals ( $b = 0.5561$ ), but I predicted that

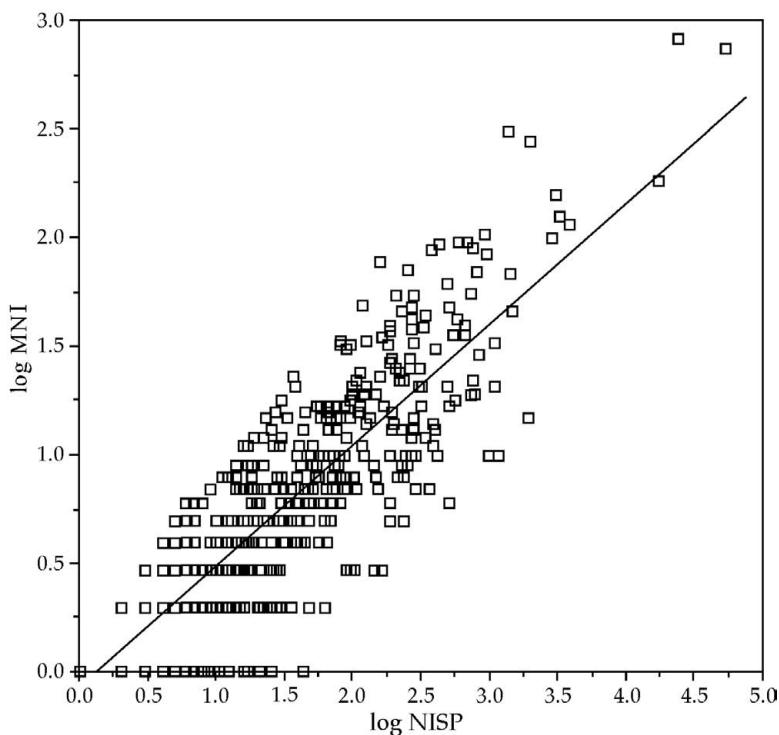


FIGURE 8.4. Relationship between NISP and MNI in thirty-five archaeological assemblages of mammal remains from North America. The number of data points is 764; not all are visible because of duplication and overlap.

the line for the latter would be less steep due to greater fragmentation. Perhaps even more bizarre is the fact that the slope of the line for the paleontological birds ( $b = 0.483$ ) is less steep than that for archaeological birds ( $b = 0.629$ ), suggesting that to contribute another MNI the NISP of paleontological birds must increase more than the NISP of archaeological birds must increase.

I am thwarted in my effort to understand why the various sets of NISP–MNI data define the relationships that they do. This is so in part because I lack data on fragmentation intensity and on which skeletal elements were identified for each taxon. It is also important to note that (i) the relationship between NISP and MNI always approximates the model described in Figure 2.4, (ii) for any given assemblage the relationship between NISP and MNI likely will be particularistic because it is historically contingent (how many skeletal elements are broken and contribute more than one NISP, and how many skeletal elements of one carcass of each taxon are identified), and (iii) differences between the relationship of the two variables across multiple assemblages may reveal something about the distinct nature of the

assemblages. Some may be more intensively fragmented, some may have been more thoroughly identified, and so on. Most importantly in the context of this volume, we have learned a bit about what kinds of data are required to begin to account for how NISP and MNI are related in any given instance.

The exploratory analysis reinforces a point I have tried to make throughout the volume. That point simply is: be explicit in your identification of a target variable, and take into account how a measured variable might, or might not, reflect the magnitude of that target variable. Thinking about the latter likely will prompt you to record data that you might not otherwise have recorded. In the case of Figures 8.1–8.4, those data might well be fragmentation (NISP:MNE ratios), determination of how many skeletal elements are identifiable in one complete skeleton, some other variable, or some combination of these. And that, it seems to me, is a good reason to know about quantifying paleofaunal remains.



# GLOSSARY

---

**absolute frequency** A raw tally or count of entities or phenomena (see **relative frequency**).

**accuracy** Correctness or exactness; the degree to which a measure conforms to the true value (an estimate is less accurate than a measurement).

**assemblage** The entire set of faunal remains from a specified context; the context may be arbitrarily, archaeologically, geologically, or biologically defined or defined in some other way (synonym: *collection*).

**biocoenose** A living community of organisms.

**closed array** Quantities are given as proportions or percentages and thus must sum to 1.0 (for proportions) or 100 percent, respectively.

**community** A set of organisms that live together and together form a more or less discrete entity; organisms comprising a community may, or may not, be functionally interlinked through competition or some other process (see Chapter 2).

**continuous variable** A variable that can take any value in a series and for which there is yet another value intermediate between any two values.

**death assemblage** See **thanatocoenose**.

**derived measurement** A measurement based on multiple fundamental measurements, such as a ratio of length to width.

**discontinuous variable** A variable for which it is possible to find two values between which there is no intermediate value.

**distal community** One or more biological communities from which remains of animals originated and which are some greater or lesser distance from the location from which the remains were collected (after Shotwell 1955, 1958).

**diversity** A general term concerning any of several variables either individually or in combination; *alpha diversity*, *beta diversity*, *gamma diversity*.

**element** See **skeletal element**.

**estimate** A description, perhaps a value, assigned to a phenomenon based on incomplete information (less accurate than a measurement).

**estimation** The act of making an estimate.

**faunule** An assemblage of associated animal remains recovered from one or several contiguous strata and dominated by members of one biological community (Tedford 1970:677).

**flat measurement** A complex measurement that is conceptual or abstract and not easily observed (synonym: *proxy measurement*).

**fidelity studies** Actualistic (experimental, ethnoarchaeological, neotaphonomic) research aimed at determining how well a future fossil record reflects the quantitative characteristics of a biological community in terms of any chosen biological variable, including morphological classes, age classes, taxonomic richness, taxonomic abundance, and trophic structure.

**fundamental measurement** A measurement that describes an easily observed property or characteristic, such as length or width (see **derived measurement** and **flat measurement**).

**identified assemblage** The set of faunal remains identified to taxon and studied by the paleozoologist, typically a fraction of the taphocoenose.

**interval scale** Measures greater than, less than relationships, and how much (distances between any two values are known), and has an arbitrary zero.

**local fauna** A set of faunal remains from one locality or several closely grouped localities that are stratigraphically equivalent or nearly so, thus the represented taxa are close in space and time (Tedford 1970:678).

**measured variable** The variable that is measured (see **target variable**).

**measurement** Writing descriptions of phenomena according to rules; specifically, the act of assigning a numerical value to an observation based on some rule(s) of assignment (see **derived measurement**, **flat measurement**, **fundamental measurement**, and **proxy measurement**).

**MNI** Minimum number of individuals (see Table 2.4).

**NISP** Number of identified specimens.

**nominal scale** Measures differences in kind, not magnitude; measurements of this scale are sometimes referred to as *qualitative attributes* or *discontinuous variables*.

**ordinal scale** Measures greater than, less than relationships, but not how much.

**proximal community** The biological community from which remains of animals originated and which is essentially geographically coincident with the location from which the remains were collected (after Shotwell 1955, 1958).

**proxy measurement** See **flat measurement**.

**quantitative variable** Variables measured on interval scales and ratio scales.

**rank order** Arrangement of a set of phenomena in a series from greatest to least magnitude, or least to greatest magnitude, but in which the distance in between any pair of phenomena is unknown.

**ratio scale** Measures greater than, less than relationships, and how much (the distances between any two values are known), and has a natural zero.

**relative frequency** A quantity or estimate that is stated in terms of another quantity or estimate (see **absolute frequency** and **closed array**).

**reliability** Replicability; repeatability; measuring something twice and obtaining the same answer.

**skeletal element** A complete, discrete anatomical unit or organ, such as a bone, tooth, or shell.

**skeletal part** Same as **specimen** but sometimes used in this book to denote a less inclusive and more restricted category, such as denoting only specimens of humerus; a synonym used in this book is *skeletal portion*.

**specimen** A bone, tooth, or shell or fragment thereof.

**taphocoenose** The set of remains of organisms with a geological mode of occurrence and found spatially and geologically associated; may be a fraction of a thanatocoenose.

**taphonomy** “The study of the transition (in all details) of animal remains from the biosphere to the lithosphere” (Efremov 1940:85).

**target variable** The variable one is interested in and seeks to measure or estimate (see **measured variable**).

**thanatocoenose** A set (assemblage) of dead organisms (synonym: *death assemblage*); may be a fraction of a biocoenose.

**validity** Measurement of an attribute that reflects the concept that we wish to describe; measuring the variable of interest rather than another variable.

**variable** A property or characteristic that can take on different values or magnitudes.



## REFERENCES

---

- Abe, Y., C. W. Marean, P. J. Nilssen, Z. Assefa, and E. C. Stone. 2002. The Analysis of Cutmarks on Archaeofauna: A Review and Critique of Quantification Procedures, and a New Image-Analysis GIS Approach. *American Antiquity* 67:643–663.
- Adams, B. J., and L. W. Konigsberg. 2004. Estimation of the Most Likely Number of Individuals from Commingled Human Skeletal Remains. *American Journal of Physical Anthropology* 125:138–151.
- Adams, W. R. 1949. *Faunal Remains from the Angel Site*. Unpublished Master's thesis, Department of Anthropology, Indiana University, Bloomington.
- Allen, J., and J. B. M. Guy. 1984. Optimal Estimations of Individuals in Archaeological Faunal Assemblages: How Minimal is the MNI? *Archaeology in Oceania* 19:41–47.
- Alroy, J. 2000. New Methods for Quantifying Macroevolutionary Patterns and Processes. *Paleobiology* 26:707–733.
- Ames, K. M. 1996. Life in the Big House: Household Labor and Dwelling Size on the Northwest Coast. In *People Who Lived in Big Houses: Archaeological Perspectives on Large Domestic Structures*, edited by G. Coupland and E. B. Banning, pp. 131–150. Monographs in World Archaeology No. 27. Prehistory Press, Madison, Wisconsin.
- Ames, K. M., and H. D. G. Maschner. 1999. *Peoples of the Northwest Coast: Their Archaeology and Prehistory*. Thames and Hudson, London.
- Ames, K. M., D. F. Raetz, S. Hamilton, and C. McAfee. 1992. Household Archaeology of a Southern Northwest Coast Plank House. *Journal of Field Archaeology* 19:275–290.
- Ames, K. M., C. M. Smith, W. L. Cornett, E. A. Sobel, S. C. Hamilton, J. Wolf, and D. Raetz. 1999. *Archaeological Investigations at 45 CL1 Cathlapotle (1991–1996), Ridgefield National Wildlife Refuge, Clark County, Washington: A Preliminary Report*. U.S.D.I. Fish and Wildlife Service, Region 1, Cultural Resource Series No. 13. Portland, Oregon.
- Amstrup, S. C., T. L. McDonald, and B. F. J. Manly (editors). 2006. *Handbook of Capture–Recapture Analysis*. Princeton University Press, Princeton, New Jersey.
- Andrews, P. 1990. *Owls, Caves and Fossils*. University of Chicago Press, Chicago.
- Andrews, P. 1996. Palaeoecology and Hominoid Palaeoenvironments. *Biological Reviews* 71:257–300.

- Anyonge, W. 1993. Body Mass in Large Extant and Extinct Carnivores. *Journal of Zoology* 231:339–350.
- Anyonge, W., and C. Roman. 2006. New Body Mass Estimates for *Canis dirus*, the Extinct Pleistocene Dire Wolf. *Journal of Vertebrate Paleontology* 26:209–212.
- Atmar, W., and B. D. Patterson. 1993. The Measure of Order and Disorder in the Distribution of Species in Fragmented Habitat. *Oecologia* 96:373–382.
- Avery, D. M. 1991. Micromammals, Owls and Vegetation Change in the Eastern Cape Midlands, South Africa, During the Last Millennium. *Journal of Arid Environments* 20:357–369.
- Avery, D. M. 1992. Micromammals and the Environment of Early Pastoralists at Spoeg River, Western Cape Province, South Africa. *South African Archaeological Bulletin* 47:116–121.
- Avery, D. M. 2002. Taphonomy of Micromammals from Cave Deposits at Kabwe (Broken Hill) and Twin Rivers in Central Zambia. *Journal of Archaeological Science* 29:537–544.
- Badgley, C. 1986. Counting Individuals in Mammalian Fossil Assemblages from Fluvial Environments. *Palaios* 1:328–338.
- Bambach, R. K. 1993. Seafood Through Time: Changes in Biomass, Energetics, and Productivity in the Marine Ecosystem. *Paleobiology* 19:372–397.
- Barnosky, A. D., M. A. Carrasco, and E. B. Davis. 2005. The Impact of the Species–Area Relationship on Estimates of Paleodiversity. *PLoS Biology* 3:1356–1361.
- Barrett, J. H. 1993. Bone Weight, Meat Yield Estimates and Cod (*Gadus morhua*): A Preliminary Study of the Weight Method. *International Journal of Osteoarchaeology* 3:1–18.
- Bartram, L. E., Jr., E. M. Kroll, and H. T. Bunn. 1991. Variability in Camp Structure and Bone Food Refuse Patterning at Kua San Hunter–Gatherer Camps. In *The Interpretation of Archaeological Spatial Patterning*, edited by E. M. Kroll and T. D. Price, pp. 77–148. Plenum Press, New York.
- Bartram, L. E., Jr., and C. W. Marean. 1999. Explaining the “Klasies Pattern”: Kua Ethnoarchaeology, the Die Kelders Middle Stone Age Archaeofauna, Long Bone Fragmentation and Carnivore Ravaging. *Journal of Archaeological Science* 26:9–29.
- Baxter, M. 2003. *Statistics in Archaeology*. Arnold, London.
- Bayham, F. E. 1979. Factors Influencing the Archaic Pattern of Animal Exploitation. *Kiva* 44:219–235.
- Behrensmeyer, A. K. 1975. The Taphonomy and Paleoecology of Plio-Pleistocene Vertebrate Assemblages East of Lake Rudolf, Kenya. *Bulletin of the Museum of Comparative Zoology* 146:473–578. Harvard University, Cambridge, Massachusetts.
- Behrensmeyer, A. K. 1978. Taphonomic and Ecologic Information from Bone Weathering. *Paleobiology* 4:150–162.
- Behrensmeyer, A. K., S. M. Kidwell, and R. A. Gastaldo. 2000. Taphonomy and Paleobiology. In *Deep Time: Paleobiology’s Perspective*, edited by D. H. Erwin and S. W. Wing, pp. 103–147. *Paleobiology* (Supplement) 26. Paleontological Society, Lawrence, Kansas.
- Bennington, J. B., and R. K. Bambach. 1996. Statistical Testing for Paleocommunity Recurrence: Are Similar Fossil Assemblages Ever the Same? *Palaeogeography, Palaeoclimatology, Palaeoecology* 127:107–133.

- Betts, M. W. 2000. Augmenting Faunal Quantification Procedures Through the Incorporation of Historical Documentary Evidence: An Investigation of Faunal Remains from Fort George. *Ontario Archaeology* 69:19–38.
- Binford, L. R. 1978. *Nunamiut Ethnoarchaeology*. Academic Press, New York.
- Binford, L. R. 1981. *Bones: Ancient Men and Modern Myths*. Academic Press, New York.
- Binford, L. R. 1984. *Faunal Remains from Klasies River Mouth*. Academic Press, New York.
- Binford, L. R. 1986. Comment. *Current Anthropology* 27:444–446.
- Binford, L. R. 1988. Fact and Fiction About the *Zinjanthropus* Floor: Data, Arguments, and Interpretations. *Current Anthropology* 29:123–135.
- Binford, L. R., and J. B. Bertram. 1977. Bone Frequencies – and Attritional Processes. In *For Theory Building in Archaeology*, edited by L. R. Binford, pp. 77–153. Academic Press, New York.
- Blackburn, T. M., K. J. Gaston, and N. Loder. 1999. Geographic Gradients in Body Size: A Clarification of Bergmann's Rule. *Diversity and Distributions* 5:165–174.
- Blalock, H. M. 1960. *Social Statistics*. McGraw Hill, New York.
- Blumenschine, R. J. 1986. *Early Hominid Scavenging Opportunities: Implications of Carcass Availability in the Serengeti and Ngorongoro Ecosystems*. British Archaeological Reports International Series 283. Oxford.
- Blumenschine, R. J. 1995. Percussion Marks, Tooth Marks, and Experimental Determinations of the Timing of Hominid and Carnivore Access to Long Bones at FLK *Zinjanthropus*, Olduvai Gorge, Tanzania. *Journal of Human Evolution* 29:21–51.
- Blumenschine, R. J., C. W. Marean, and S. D. Capaldo. 1996. Blind Tests of Inter-Analyst Correspondence and Accuracy in the Identification of Cut Marks, Percussion Marks, and Carnivore Tooth Marks on Bone Surfaces. *Journal of Archaeological Science* 23:493–507.
- Blumenschine, R. J., and M. M. Selvaggio. 1988. Percussion Marks on Bone Surfaces as a New Diagnostic of Hominid Behavior. *Nature* 333:763–765.
- Blumenschine, R. J., and M. M. Selvaggio. 1991. On the Marks of Marrow Bone Processing by Hammerstones and Hyaenas: Their Anatomical Patterning and Archaeological Implications. In *Cultural Beginnings: Approaches to Understanding Early Hominid Life-ways in the African Savanna*, edited by J. D. Clark, pp. 17–32. Union Internationale des Sciences Préhistoriques et Protohistoriques Monographien Band 19. Bonn.
- Bobrowsky, P. T. 1982. An Examination of Casteel's MNI Behavior Analysis: A Reductionist Approach. *Midcontinental Journal of Archaeology* 7:173–184.
- Bökonyi, S. 1970. A New Method for the Determination of the Number of Individuals in Animal Bone Material. *American Journal of Archaeology* 74:291–292.
- Brain, C. K. 1967. Hottentot Food Remains and Their Bearing on the Interpretation of Fossil Bone Assemblages. *Scientific Papers of the Namib Desert Research Station* 32:1–11.
- Brain, C. K. 1969. The Contribution of Namib Desert Hottentots to an Understanding of Australopithecine Bone Accumulations. *Scientific Papers of the Namib Desert Research Station* 39:13–22.
- Brain, C. K. 1974. Some Suggested Procedures in the Analysis of Bone Accumulations from Southern African Quaternary Sites. *Annals of the Transvaal Museum* 29:1–8.

- Brain, C. K. 1976. Some Principles in the Interpretation of Bone Accumulations Associated with Man. In *Human Origins: Louis Leakey and the East African Evidence*, edited by G. L. Isaac and E. R. McCown, pp. 97–116. W. B. Benjamin, Menlo Park, California.
- Brain, C. K. 1981. *The Hunters or the Hunted? An Introduction to African Cave Taphonomy*. University of Chicago Press, Chicago.
- Breitburg, E. 1991. Verification and Reliability of NISP and MNI Methods of Quantifying Taxonomic Abundance: A View from Historic Site Zooarchaeology. In *Beamers, Bobwhites, and Blue-Points: Tributes to the Career of Paul W. Parmalee*, edited by J. R. Purdue, W. E. Klipper, and B. W. Styles, pp. 153–162. Illinois State Museum Scientific Papers Vol. 23. Springfield.
- Broughton, J. M. 1994a. Declines in Mammalian Foraging Efficiency During the Late Holocene, San Francisco Bay, California. *Journal of Anthropological Archaeology* 13:371–401.
- Broughton, J. M. 1994b. Late Holocene Resource Intensification in the Sacramento Valley, California: The Vertebrate Evidence. *Journal of Archaeological Science* 21:501–514.
- Broughton, J. M. 1999. *Resource Depression and Intensification During the Late Holocene, San Francisco Bay: Evidence from the Emeryville Shellmound Vertebrate Fauna*. Anthropological Records Vol. 32. University of California, Berkeley.
- Broughton, J. M. 2004. *Prehistoric Human Impacts on California Birds: Evidence from the Emeryville Shellmound Avifauna*. Ornithological Monographs No. 56.
- Broughton, J. M., and D. K. Grayson. 1993. Diet Breadth, Adaptive Change, and the White Mountains Fauna. *Journal of Archaeological Science* 20:331–336.
- Broughton, J. M., V. I. Cannon, S. Arnold, R. J. Bogiatto, and K. Dalton. 2006. The Taphonomy of Owl-Deposited Fish Remains and the Origin of the Homestead Cave Ichthyofauna. *Journal of Taphonomy* 4:69–95.
- Brown, E. R. 1961. *The Black-Tailed Deer of Western Washington*. Washington State Game Department, Biological Bulletin No. 13. Olympia, Washington.
- Brown, J. H., and A. C. Gibson. 1983. *Biogeography*. C. V. Mosby, St. Louis.
- Brown, J. H., and M. V. Lomolino. 1998. *Biogeography*, 2nd ed. Sinauer Associates, Sunderland, Massachusetts.
- Bunn, H. T. 1982. *Meat-Eating and Human Evolution: Studies on the Diet and Subsistence Patterns of Plio-Pleistocene Hominids in East Africa*. Ph.D. dissertation, University of California, Berkeley.
- Bunn, H. T. 1986. Patterns of Skeletal Element Representation and Hominid Subsistence Activities at Olduvai Gorge, Tanzania, and Koobi Fora, Kenya. *Journal of Human Evolution* 15:673–690.
- Bunn, H. T. 1991. A Taphonomic Perspective on the Archaeology of Human Origins. *Annual Review of Anthropology* 20:433–467.
- Bunn, H. T. 2001. Hunting, Power Scavenging, and Butchering by Hadza Foragers and by Plio-Pleistocene Homo. In *Meat-Eating and Human Evolution*, edited by C. B. Stanford and H. T. Bunn, pp. 199–218. Oxford University Press, Oxford.
- Bunn, H. T., and E. M. Kroll. 1986. Systematic Butchery by Plio/Pleistocene Hominids at Olduvai Gorge, Tanzania. *Current Anthropology* 27:431–452.

- Bunn, H. T., and E. M. Kroll. 1988. Reply (to Binford). *Current Anthropology* 29:135–149.
- Bush, A. M., M. J. Markey, and C. R. Marshall. 2004. Removing Bias from Diversity Curves: The Effects of Spatially Organized Biodiversity on Sampling-Standardization. *Paleobiology* 30:666–686.
- Butler, V. L. 2000. Resource Depression on the Northwest Coast of North America. *Antiquity* 74:649–661.
- Butler, V. L. 2001. Changing Fish Use on Mangaia, Southern Cook Islands: Resource Depression and the Prey Choice Model. *International Journal of Osteoarchaeology* 11:88–100.
- Butler, V. L., and S. K. Campbell. 2004. Resource Intensification and Resource Depression in the Pacific Northwest of North America: A Zooarchaeological Review. *Journal of World Prehistory* 18:327–405.
- Butler, V. L., and M. G. Delacorte. 2004. Doing Zooarchaeology as if It Mattered: Use of Faunal Data to Address Current Issues in Fish Conservation Biology in Owens Valley, California. In *Zooarchaeology and Conservation Biology*, edited by R. L. Lyman and K. P. Cannon, pp. 25–44. University of Utah Press, Salt Lake City.
- Buzas, M. A., and L.-A. Hayek. 2005. On Richness and Evenness Within and Between Communities. *Paleobiology* 31:199–220.
- Byrd, J. E. 1997. The Analysis of Diversity in Archaeological Faunal Assemblages: Complexity and Subsistence Strategies in the Southeast During the Middle Woodland Period. *Journal of Anthropological Archaeology* 16:49–72.
- Cain, C. R. 2005. Using Burned Animal Bone to Look at Middle Stone Age Occupation and Behavior. *Journal of Archaeological Science* 32:873–884.
- Cain, S. A. 1938. The Species Area Curve. *American Midland Naturalist* 19:573–581.
- Cannon, M. D. 1999. A Mathematical Model of the Effects of Screen Size on Zooarchaeological Relative Abundance Measures. *Journal of Archaeological Science* 26:205–214.
- Cannon, M. D. 2000. Large Mammal Relative Abundance in Pithouse and Pueblo Period Archaeofaunas from Southwestern New Mexico: Resource Depression Among the Mimbres–Mogollon? *Journal of Anthropological Archaeology* 19:317–347.
- Cannon, M. D. 2001. Archaeofaunal Relative Abundance, Sample Size, and Statistical Methods. *Journal of Archaeological Science* 28:185–195.
- Cannon, M. D. 2003. A Model of Central Place Forager Prey Choice and an Application to Faunal Remains from the Mimbres Valley, New Mexico. *Journal of Anthropological Archaeology* 22:1–25.
- Capaldo, S. D. 1997. Experimental Determinations of Carcass Processing by Plio-Pleistocene Hominids and Carnivores at FLK 22 (*Zinjanthropus*), Olduvai Gorge, Tanzania. *Journal of Human Evolution* 33:555–597.
- Capaldo, S. D., and R. J. Blumenschine. 1994. A Quantitative Diagnosis of Notches Made by Hammerstone Percussion and Carnivore Gnawing in Bovid Long Bones. *American Antiquity* 59:724–748.
- Carder, N., E. J. Reitz, and J. M. Compton. 2004. Animal Use in the Georgia Pine Barrens: An Example from the Hartford Site (9PU1). *Southeastern Archaeology* 24:25–40.

- Carmines, E. G., and R. Z. Zeller. 1979. *Reliability and Validity Assessment*. Sage University Paper 17. Beverly Hills, California.
- Casteel, R. W. 1972. Some Biases in the Recovery of Archaeological Faunal Remains. *Proceedings of the Prehistoric Society* 38:382–388.
- Casteel, R. W. 1974. A Method for Estimation of Live Weight of Fish from the Size of Skeletal Remains. *American Antiquity* 39:94–97.
- Casteel, R. W. 1977. A Consideration of the Behavior of the Minimum Number of Individuals Index: A Problem in Faunal Characterization. *Ossa* 3/4:141–151.
- Casteel, R. W. 1978. Faunal Assemblages and the “Wiegemethode” or Weight Method. *Journal of Field Archaeology* 5:71–77.
- Casteel, R. W. n.d. A Treatise on the Minimum Number of Individuals Index: An Analysis of its Behaviour and a Method for Its Prediction. Unpublished manuscript.
- Casteel, R. W., and D. K. Grayson. 1977. Terminological Problems in Quantitative Faunal Analysis. *World Archaeology* 9:235–242.
- Chao, A., R. L. Chazdon, R. K. Colwell, and T. Shen. 2005. A New Statistical Approach for Assessing Similarity of Species Composition with Incidence and Abundance Data. *Ecology Letters* 8:148–159.
- Chaplin, R. E. 1971. *The Study of Animal Bones from Archaeological Sites*. Seminar Press, London.
- Cheetham, A. H., and J. E. Hazel. 1969. Binary (Presence–Absence) Similarity Coefficients. *Journal of Paleontology* 43:1130–1136.
- Claassen, C. 1998. *Shells*. Cambridge University Press, Cambridge, UK.
- Clark, J., and T. E. Guensburg. 1970. Population Dynamics of *Leptomyers*. *Fieldiana: Geology* 16:411–451.
- Clason, A. T. 1972. Some Remarks on the Use and Presentation of Archaeozoological Data. *Helinium* 12:139–153.
- Clason, A. T., and W. Prummel. 1977. Collecting, Sieving and Archaeozoological Research. *Journal of Archaeological Science* 4:171–175.
- Cleghorn, N., and C. W. Marean. 2004. Distinguishing Selective Transport and *In Situ* Attrition: A Critical Review of Analytical Approaches. *Journal of Taphonomy* 2:43–67.
- Cleland, C. E. 1966. *The Prehistoric Animal Ecology and Ethnozoology of the Upper Great Lakes Region*. Anthropological Papers No. 29. Museum of Anthropology, University of Michigan, Ann Arbor.
- Cleland, C. E. 1976. The Focal–Diffuse Model: An Evolutionary Perspective on the Prehistoric Cultural Adaptations of the Eastern United States. *Mid-Continental Journal of Archaeology* 1:59–76.
- Colwell, R. K., and J. A. Coddington. 1994. Estimating Terrestrial Biodiversity Through Extrapolation. *Philosophical Transactions of the Royal Society of London B* 345:101–118.
- Colwell, R. K., C. X. Mao, and J. Chang. 2004. Interpolating, Extrapolating, and Comparing Incidence-Based Species Accumulation Curves. *Ecology* 85:2717–2727.
- Cook, S. F., and A. E. Treganza. 1950. The Quantitative Investigation of Indian Mounds with Special Reference to the Relation of the Physical Components to the Probable Material

- Culture. *University of California Publications in American Archaeology and Ethnology* 40:223–262.
- Cooper, R. A., P. A. Maxwell, J. S. Crampton, A. G. Beau, C. M. Jones, and B. A. Marshall. 2006. Completeness of the Fossil Record: Estimating Losses Due to Small Body Size. *Geology* 34:241–244.
- Crader, D. C. 1983. Recent Single-Carcass Bone Scatters and the Problem of “Butchery” Sites in the Archaeological Record. In *Animals and Archaeology: I. Hunters and Their Prey*, edited by J. Clutton-Brock and C. Grigson, pp. 107–141. British Archaeological Reports, International Series 163. Oxford.
- Crampton, J. S., A. G. Geu, R. A. Cooper, C. M. Jones, B. Marshall, and P. A. Maxwell. 2003. Estimating the Rock Volume Bias in Paleobiodiversity Studies. *Science* 301:358–360.
- Cruz-Uribe, K., and R. G. Klein. 1994. Chew Marks and Cut Marks on Animal Bones from the Kasteelberg B and Dune Field Midden Later Stone Age Sties, Western Cape Province, South Africa. *Journal of Archaeological Science* 21:35–49.
- Cutler, A. 1991. Nested Faunas and Extinction in Fragmented Habitats. *Conservation Biology* 5:496–505.
- Cutler, A. 1994. Nested Biotas and Biological Conservation: Metrics, Mechanisms, and Meaning of Nestedness. *Landscape and Urban Planning* 28:73–82.
- Damuth, J. 1982. Analysis of the Preservation of Community Structure in Assemblages of Fossil Mammals. *Paleobiology* 8:434–446.
- Damuth, J., and B. J. MacFadden (editors). 1990. *Body Size in Mammalian Paleobiology: Estimation and Biological Implications*. Cambridge University Press, Cambridge, UK.
- Darwent, C., and R. L. Lyman. 2002. Detecting the Postburial Fragmentation of Carpal, Tarsal, and Phalanges. In *Advances in Forensic Taphonomy*, edited by W. D. Haglund and M. H. Sorg, pp. 355–377. CRC Press, Boca Raton, Florida.
- Dean, R. M. 2001. Social Change and Hunting During the Pueblo III to Pueblo IV Transition, East-Central Arizona. *Journal of Field Archaeology* 28:271–285.
- Dean, R. M. 2005a. Site-Use Intensity, Cultural Modification of the Environment, and the Development of Agricultural Communities in Southern Arizona. *American Antiquity* 70:403–431.
- Dean, R. M. 2005b. Old Bones: The Effects of Curation and Exchange on the Interpretation of Artiodactyl Remains in Hohokam Sites. *Kiva* 70:255–272.
- Dechert, B. 1995. The Bone Remains from Hirbet-Ez Zeraqon. In *Archaeozoology of the Near East II*, edited by H. Buitenhuis and H.-P. Uerpmann, pp. 79–87. Backhuys, Leiden, The Netherlands.
- Dodson, P. 1973. The Significance of Small Bones in Paleoecological Interpretation. *University of Wyoming Contributions in Geology* 12:15–19.
- Dodson, P., and D. Wexlar. 1979. Taphonomic Investigations of Owl Pellets. *Paleobiology* 5:275–284.
- Domínguez-Rodrigo, M. 1997. Meat-Eating by Early Hominids at the FLK 22 *Zinjanthropus* Site, Olduvai Gorge (Tanzania): An Experimental Approach Using Cut-Mark Data. *Journal of Human Evolution* 33:669–690.

- Domínguez-Rodrigo, M. 1999a. Flesh Availability and Bone Modifications in Carcasses Consumed by Lions: Palaeoecological Relevance in Hominid Foraging Patterns. *Palaeogeography, Palaeoclimatology, Palaeoecology* 149:373–388.
- Domínguez-Rodrigo, M. 1999b. Distinguishing Between Apples and Oranges: The Application of Modern Cut-Mark Studies to the Plio-Pleistocene (A Reply to Monahan). *Journal of Human Evolution* 37:793–800.
- Domínguez-Rodrigo, M. 2002. Hunting and Scavenging by Early Humans: The State of the Debate. *Journal of World Prehistory* 16:1–54.
- Domínguez-Rodrigo, M. 2003a. Bone Surface Modifications, Power Scavenging and the “Display” Model at Early Archaeological Sites: A Critical Review. *Journal of Human Evolution* 45:411–416.
- Domínguez-Rodrigo, M. 2003b. On Cut Marks and Statistical Inferences: Methodological Comments on Lupo & O’Connell (2002). *Journal of Archaeological Science* 30:381–386.
- Domínguez-Rodrigo, M., and R. Barba. 2006. New Estimates of Tooth Mark and Percussion Mark Frequencies at the FLK Zinj Site: The Carnivore–Hominid–Carnivore Hypothesis Falsified. *Journal of Human Evolution* 50:170–194.
- Domínguez-Rodrigo, M., and T. R. Pickering. 2003. Early Hominid Hunting and Scavenging: A Zooarchaeological Review. *Evolutionary Anthropology* 12:275–282.
- Driver, J. C. 1992. Identification, Classification and Zooarchaeology. *Circaeia* 9(1):35–47.
- Ducos, P. 1968. *L’Origine des Animaux Domestiques en Palestine*. Publications de l’Institut de Préhistoire de l’Université de Bordeaux Mémoire 6.
- Dunnell, R. C. 1967. *Prehistory of Fishtrap Kentucky: Archaeological Interpretation in Marginal Areas*. Doctoral dissertation, Department of Anthropology, Yale University, New Haven, Connecticut.
- Dunnell, R. C. 1972. *Prehistory of Fishtrap Kentucky*. Yale University Publications in Anthropology No. 75. New Haven, Connecticut.
- During, E. 1986. The Fauna of Alvastra: An Osteological Analysis of Animal Bones from a Neolithic Pile Dwelling. *Ossa* 12(Supplement 1):1–210.
- Dyck, I., and R. E. Morlan. 1995. *The Sjovold Site: A River Crossing Campsite in the Northern Plains*. Mercury Series, Archaeological Survey of Canada, Paper 151. Canadian Museum of Civilization, Hull, Quebec.
- Edgar, H. J. H., and P. W. Sciumelli. 2006. Comparative Human and Deer (*Odocoileus virginianus*) Taphonomy at the Richards Site, Ohio. *International Journal of Osteoarchaeology* 16:124–137.
- Efremov, I. A. 1940. Taphonomy: New Branch of Paleontology. *Pan-American Geologist* 74:81–93.
- Egeland, C. P. 2003. Carcass Processing Intensity and Cutmark Creation: An Experimental Approach. *Plains Anthropologist* 48:39–51.
- Egi, N. 2001. Body Mass Estimates in Extinct Mammals from Limb Bone Dimensions: The Case of North American Hyaenodontids. *Paleontology* 44:497–528.
- Ellis, D. V. n.d. Untitled manuscript report on the 1984 Willamette Associates’ investigations in the Meier site locality. Unpublished report on file, Department of Anthropology, Portland State University, Portland.

- Emerson, T. E. 1978. A New Method for Calculating the Live Weight of the Northern White-tailed Deer from Osteoarchaeological Material. *Midcontinental Journal of Archaeology* 3:35–44.
- Emerson, T. E. 1983. From Bones to Venison: Calculating the Edible Meat of a White-Tailed Deer. In *Prairie Archaeology*, edited by G. E. Gibbon, pp. 63–73. University of Minnesota Publications in Anthropology No. 3. Minneapolis.
- Enloe, J. G. 2003a. Acquisition and Processing of Reindeer in the Paris Basin. In *Zooarchaeological Insights into Magdalenian Lifeways*, edited by S. Costamagno and V. Laroulandie, pp. 23–31. British Archaeological Reports International Series 1144. Oxford.
- Enloe, J. G. 2003b. Food Sharing Past and Present: Archaeological Evidence for Economic and Social Interactions. *Before Farming: The Archaeology and Anthropology of Hunter-Gatherers* 1:1–23.
- Enloe, J. G., and F. David. 1992. Food Sharing in the Paleolithic: Carcass Refitting at Pincevent. In *Piecing Together the Past: Applications of Refitting Studies in Archaeology*, edited by J. L. Hoffman and J. G. Enloe, pp. 296–315. British Archaeological Reports International Series 578. Oxford.
- Enloe, J. G., F. David, and G. Baryshnikov. 2000. Hyenas and Hunters: Zooarchaeological Investigations at Prolov II Cave, Crimea. *International Journal of Osteoarchaeology* 10:310–324.
- Errington, P. L. 1930. The Pellet Analysis Method of Raptor Food Habits Study. *Condor* 32:292–296.
- Everitt, B. S. 1977. *The Analysis of Contingency Tables*. Chapman and Hall, London.
- Ewen, C. R. 1986. Fur Trade Archaeology: A Study of Frontier Hierarchies. *Historical Archaeology* 20:15–28.
- Fagerstrom, J. A. 1964. Fossil Communities in Paleoecology: Their Recognition and Significance. *Geological Society of America Bulletin* 75:1197–1216.
- Faith, J. T., and A. D. Gordon. 2007. Skeletal Element Abundances in Archaeofaunal Assemblages: Economic Utility, Sample Size, and Assessment of Carcass Transport Strategies. *Journal of Archaeological Science* 34:872–882.
- Fernandez-Jalvo, Y., and P. Andrews. 1992. Small Mammal Taphonomy of Gran Dolina, Atapuerca (Burgos), Spain. *Journal of Archaeological Science* 19:407–428.
- Fieller, N. R. J., and A. Turner. 1982. Number Estimation in Vertebrate Samples. *Journal of Archaeological Science* 9:49–62.
- Fisher, A. K. 1896. Food of the Barn Owl (*Strix pratincola*). *Science* 3:624–625.
- Fisher, J. W., Jr. 1995. Bone Surface Modifications in Zooarchaeology. *Journal of Archaeological Method and Theory* 2:7–68.
- Fisher, R. A., A. S. Corbet, and C. B. Williams. 1943. The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population. *Journal of Animal Ecology* 12:42–58.
- Flannery, K. V. 1965. The Ecology of Early Food Production in Mesopotamia. *Science* 147:1247–1256.

- Flannery, K. V. 1969. Origins and Ecological Effects of Early Domestication in Iran and the Near East. In *The Domestication and Exploitation of Plants and Animals*, edited by P. J. Ucko and G. W. Dimbleby, pp. 73–100. Aldine, Chicago.
- Ford, J. A. 1962. *A Quantitative Method for Deriving Cultural Chronology*. Pan American Union, Technical Bulletin No. 1.
- Francillon-Vieillot, H., V. de Buffrénil, J. Castanet, J. Géraudi, F. J. Meunier, J. Y. Sire, L. Zylberberg, and A. de Ricqlès. 1990. Microstructure and Mineralization of Vertebrate Skeletal Tissues. In *Skeletal Biominerization: Patterns, Processes, and Evolutionary Trends*, edited by J. G. Carter, pp. 471–530. Van Nostrand Reinhold, New York.
- Gamble, C. 1978. Optimising Information from Studies of Faunal Remains. In *Sampling in Contemporary British Archaeology*, edited by J. F. Cherry, C. Gamble, and S. Shennan, pp. 321–353. British Archaeological Reports British Series 50. Oxford.
- Gangestad, S. W., and R. Thornhill. 1998. The Analysis of Fluctuating Asymmetry Redux: The Robustness of Parametric Statistics. *Animal Behavior* 55:497–501.
- Gargett, R. H., and D. Vale. 2005. There's Something Fishy Going on Around Here. *Journal of Archaeological Science* 32:647–652.
- Gaston, K. J. 1996. Species Richness: Measure and Measurement. In *Biodiversity: A Biology of Numbers and Difference*, edited by K. J. Gaston, pp. 77–113. Blackwell Scientific, Oxford.
- Gautier, A. 1984. How Do I Count You, Let Me Count the Ways? Problems of Archaeozoological Quantification. In *Animals and Archaeology 4: Husbandry in Europe*, edited by C. Grigson and J. Clutton-Brock, pp. 237–251. British Archaeological Reports, International Series 227. Oxford.
- Gautier, A. 1993. Trace Fossils in Archaeozoology. *Journal of Archaeological Science* 20:511–523.
- Gifford-Gonzalez, D. P. 1989. Ethnographic Analogs for Interpreting Modified Bones: Some Cases from East Africa. In *Bone Modification*, edited by R. Bonnichsen and M. H. Sorg, pp. 179–246. Center for the Study of the First Americans, Orono, Maine.
- Gifford-Gonzalez, D. P. 1991. Bones Are Not Enough: Analogues, Knowledge, and Interpretive Strategies in Zooarchaeology. *Journal of Anthropological Archaeology* 10:215–254.
- Gilbert, A. S. 1979. *Urban Taphonomy and Mammalian Remains from the Bronze Age of Godin Tepe, Western Iran*. Doctoral dissertation, Columbia University, New York. University Microfilms, Ann Arbor.
- Gilbert, A. S., and B. H. Singer. 1982. Reassessing Zooarchaeological Quantification. *World Archaeology* 14:21–40.
- Gilbert, A. S., B. H. Singer, and D. Perkins, Jr. 1981. Quantification Experiments on Computer-Simulated Faunal Collections. *Ossa* 8:79–94.
- Gilbert, B. M. 1969. Some Aspects of Diet and Butchering Techniques Among Prehistoric Indians in South Dakota. *Plains Anthropologist* 14:277–294.
- Gilinsky, N. L., and J. B. Bennington. 1994. Estimating Numbers of Whole Individuals from Collections of Body Parts: A Taphonomic Limitation of the Paleontological Record. *Paleobiology* 20:245–258.
- Gobalet, K. W. 2001. A Critique of Faunal Analysis: Inconsistency Among Experts in Blind Tests. *Journal of Archaeological Science* 28:377–386.

- Gobalet, K. W. 2005. Comment on "Size Matters: 3-mm Sieves Do Not Increase Richness in a Fishbone Assemblage from Arrawarra I, An Aboriginal Australian Shell Midden on the Mid-North Coast of New South Wales, Australia" by Vale and Gargett. *Journal of Archaeological Science* 32:643–645.
- Gotelli, N. J., and R. K. Colwell. 2001. Quantifying Biodiversity: Procedures and Pitfalls in the Measurement and Comparison of Species Richness. *Ecology Letters* 4:379–391.
- Graham, R. W. 1984. Paleoenvironmental Implications of the Quaternary Distribution of the Eastern Chipmunk (*Tamias striatus*) in Central Texas. *Quaternary Research* 21:111–114.
- Grayson, D. K. 1973. On the Methodology of Faunal Analysis. *American Antiquity* 39:432–439.
- Grayson, D. K. 1978a. Minimum Numbers and Sample Size in Vertebrate Faunal Analysis. *American Antiquity* 43:53–65.
- Grayson, D. K. 1978b. Reconstructing Mammalian Communities: A Discussion of Shotwell's Method of Paleoecological Analysis. *Paleobiology* 4:77–81.
- Grayson, D. K. 1979. On the Quantification of Vertebrate Archaeofaunas. In *Advances in Archaeological Method and Theory*, Vol. 2, edited by M. B. Schiffer, pp. 199–237. Academic Press, New York.
- Grayson, D. K. 1981a. A Critical View of the Use of Archaeological Vertebrates in Paleoenvironmental Reconstruction. *Journal of Ethnobiology* 1:28–38.
- Grayson, D. K. 1981b. The Effects of Sample Size on Some Derived Measures in Vertebrate Faunal Analysis. *Journal of Archaeological Science* 8:77–88.
- Grayson, D. K. 1984. *Quantitative Zooarchaeology: Topics in the Analysis of Archaeological Faunas*. Academic Press, Orlando, Florida.
- Grayson, D. K. 1988. *Danger Cave, Last Supper Cave, and Hanging Rock Shelter: The Faunas*. American Museum of Natural History Anthropological Papers Vol. 66, No. 1. New York.
- Grayson, D. K. 1991a. Alpine Faunas from the White Mountains, California: Adaptive Change in the Late Prehistoric Great Basin? *Journal of Archaeological Science* 18:483–506.
- Grayson, D. K. 1991b. The Small Mammals of Gatecliff Shelter: Did People Make a Difference? In *Beamers, Bobwhites, and Blue-Points: Tributes to the Career of Paul W. Parmalee*, edited by J. R. Purdue, W. E. Klippel, and B. W. Styles, pp. 99–109. Illinois State Museum Scientific Papers 23. Springfield.
- Grayson, D. K. 1998. Moisture History and Small Mammal Community Richness During the Latest Pleistocene and Holocene, Northern Bonneville Basin, Utah. *Quaternary Research* 49:330–334.
- Grayson, D. K. 2000. The Homestead Cave Mammals. In *Late Quaternary Paleoecology in the Great Basin*, edited by D. B. Madsen, pp. 67–89. Utah Geological Survey Bulletin 130. Salt Lake City.
- Grayson, D. K., and F. Delpech. 1994. The Evidence for Middle Paleolithic Scavenging from Couche VIII, Grotte Valley (Dordogne, France). *Journal of Archaeological Science* 21:359–375.
- Grayson, D. K., and F. Delpech. 1998. Changing Diet Breadth in the Early Upper Paleolithic of Southwestern France. *Journal of Archaeological Science* 25:1119–1129.

- Grayson, D. K., F. Delpech, J.-P. Rigaud, and J. F. Simek. 2001. Explaining the Development of Dietary Dominance by a Single Ungulate Taxon at Grotte XVI, Dordogne, France. *Journal of Archaeological Science* 28:115–125.
- Grayson, D. K., and C. J. Frey. 2004. Measuring Skeletal Part Representation in Archaeological Faunas. *Journal of Taphonomy* 2:27–42.
- Greenfield, H. J. 1999. The Origins of Metallurgy: Distinguishing Stone from Metal Cut-Marks on Bones from Archaeological Sites. *Journal of Archaeological Science* 26:797–808.
- Guilday, J. E. 1970. Animal Remains from Archeological Excavations at Fort Ligonier. *Annals of the Carnegie Museum* 42:177–186.
- Guilday, J. E., P. W. Parmalee, and D. P. Tanner. 1962. Aboriginal Butchering Techniques at the Eschelman Site (36LA12), Lancaster County, Pennsylvania. *Pennsylvania Archaeologist* 32(2):59–83.
- Guthrie, R. D. 1968. Paleoecology of the Large-Mammal Community in Interior Alaska During the Late Quaternary. *American Midland Naturalist* 79:346–363.
- Guthrie, R. D. 1982. Mammals of the Mammoth Steppe as Paleoenvironmental Indicators. In *Paleoecology of Beringia*, edited by D. M. Hopkins, J. V. Matthews, Jr., C. E. Schweger, and S. B. Young, pp. 307–326. Academic Press, New York.
- Guthrie, R. D. 1984a. Alaskan Megabucks, Megabulls, and Megagrams: The Issue of Pleistocene Gigantism. In *Contributions in Quaternary Vertebrate Paleontology: A Volume in Memorial to John E. Guilday*, edited by H. H. Genoways and M. R. Dawson, pp. 482–510. Carnegie Museum of Natural History Special Publication No. 8. Pittsburgh, Pennsylvania.
- Guthrie, R. D. 1984b. Mosaics, Allelochemics, and Nutrients: An Ecological Theory of Late Pleistocene Megafaunal Extinctions. In *Quaternary Extinctions: A Prehistoric Revolution*, edited by P. S. Martin and R. G. Klein, pp. 259–298. University of Arizona Press, Tucson.
- Hadly, E. A. 1999. Fidelity of Terrestrial Vertebrate Fossils to a Modern Ecosystem. *Palaeogeography, Palaeoclimatology, Palaeoecology* 149:389–409.
- Hardesty, D. L. 1975. The Niche Concept: Suggestions for Its Use in Human Ecology. *Human Ecology* 3:71–85.
- Haynes, G. 1980. Evidence of Carnivore Gnawing on Pleistocene and Recent Mammalian Bones. *Paleobiology* 6:341–351.
- Haynes, G. 1988. Mass Deaths and Serial Predation: Comparative Taphonomic Studies of Modern Large Mammal Death Sites. *Journal of Archaeological Science* 15:119–135.
- Haynes, G. 2002. Archeological Methods for Reconstructing Human Predation on Terrestrial Vertebrates. In *The Fossil Record of Predation*, edited by M. Kowalewski and P. H. Kelley, pp. 51–67. Paleontological Society Paper No. 8. New Haven, Connecticut.
- Henderson, R. A., and M. L. Heron. 1977. A Probabilistic Method of Paleobiogeographic Analysis. *Lethaia* 10:1–15.
- Hesse, B. 1982. Bias in the Zooarchaeological Record: Suggestions for Interpretation of Bone Counts in Faunal Samples from the Plains. In *Plains Indian Studies: A Collection of Essays in Honor of John C. Ewers and Waldo R. Wedel*, edited by D. H. Ubelaker and H. J. Viola, pp. 157–172. Smithsonian Contributions to Anthropology No. 30. Washington, DC.

- Hesse, B., and P. Wapnish. 1985. *Animal Bone Archeology: From Objectives to Analysis*. Manuals on Archeology 5. Taraxacum, Washington, DC.
- Hibbard, C. W. 1949. Techniques of Collecting Microvertebrate Fossils. *University of Michigan Contributions from the Museum of Paleontology* 7(2):7–19.
- Higham, C. F. W. 1968. Faunal Sampling and Economic Prehistory. *Zeitschrift für Saugertierkunde* 33:297–305.
- Hoffman, A. 1979. Community Paleoecology as an Epiphenomenal Science. *Paleobiology* 5:357–379.
- Hoffman, R. 1988. The Contribution of Raptorial Birds to Patterning in Small Mammal Assemblages. *Paleobiology* 14:81–90.
- Holland, S. 2005. *Analytical Rarefaction*, version 1.3. <http://www.uga.edu/~strata/software/software.html>. Accessed Nov. 15, 2006.
- Holtzman, R. C. 1979. Maximum Likelihood Estimation of Fossil Assemblage Composition. *Paleobiology* 5:77–90.
- Hopkins, H. L., and M. L. Kennedy. 2004. An Assessment of Indices of Relative and Absolute Abundance for Monitoring Populations of Small Mammals. *Wildlife Society Bulletin* 32:1289–1296.
- Horton, D. R. 1984. Minimum Numbers: A Consideration. *Journal of Archaeological Science* 11:255–271.
- Howard, H. 1930. A Census of the Pleistocene Birds of Rancho La Brea from the Collections of the Los Angeles Museum. *Condor* 32:81–88.
- Hudson, J. 1990. *Advancing Methods in Zooarchaeology: An Ethnoarchaeological Study Among the Aka Pygmies*. Ph.D. dissertation, Department of Anthropology, University of California, Santa Barbara.
- Hudson, R. J., J. C. Haigh, and A. B. Bubenik. 2002. Physical and Physiological Adaptations. In *North American Elk: Ecology and Management*, edited by D. E. Toweill and J. W. Thomas, pp. 199–257. Smithsonian Institution Press, Washington, DC.
- Huelsbeck, D. R. 1989. Zooarchaeological Measures Revisited. *Historical Archaeology* 23:113–117.
- Hurlbert, S. H. 1971. The Nonconcept of Species Diversity: A Critique and Alternative Parameters. *Ecology* 52:577–586.
- Jackson, H. E. 1989. The Trouble with Transformations: Effects of Sample Size and Sample Composition on Meat Weight Estimates Based on Skeletal Mass Allometry. *Journal of Archaeological Science* 16:601–610.
- Jackson, H. E., and S. L. Scott. 2002. Woodland Faunal Exploitation in the Midsouth. In *The Woodland Southeast*, edited by D. G. Anderson and R. C. Mainfort, Jr., pp. 461–482. University of Alabama Press, Tuscaloosa.
- Jackson, J. B. C., and K. G. Johnson. 2001. Measuring Past Biodiversity. *Science* 293:2401–2404.
- Jacobson, J. A. 2003. Identification of mule deer (*Odocoileus hemionus*) and white-tailed deer (*Odocoileus virginianus*) postcranial remains as a means of determining human subsistence strategies. *Plains Anthropologist* 48:287–297.

- Jacobson, J. A. 2004. Determining human ecology on the Plains through the identification of mule deer (*Odocoileus hemionus*) and white-tailed deer (*Odocoileus virginianus*) postcranial remains. Unpublished Ph.D. dissertation, University of Tennessee, Knoxville, Tennessee.
- James, S. R. 1990. Monitoring Archaeofaunal Changes During the Transition to Agriculture in the American Southwest. *Kiva* 56:25–43.
- James, S. R. 1997. Methodological Issues Concerning Screen Size Recovery Rates and Their Effects on Archaeofaunal Interpretations. *Journal of Archaeological Science* 24:385–397.
- Jamniczky, H. A., D. B. Brinkman, and A. P. Russell. 2003. Vertebrate Microsite Sampling: How Much is Enough? *Journal of Vertebrate Paleontology* 23:725–734.
- Janson, S., and J. Vegelius. 1981. Measures of Ecological Association. *Oecologia* 49:371–376.
- Johnson, E. 1989. Human Modified Bones from Early Southern Plains Sites. In *Bone Modification*, edited by R. Bonnichsen and M. H. Sorg, pp. 431–471. University of Maine Center for the Study of the First Americans, Orono.
- Johnson, R. E., and K. M. Cassidy. 1997. *Terrestrial Mammals of Washington State: Location Data and Predicted Distributions*. Vol. 3. Washington State Gap Analysis Project Final Report, Washington Cooperative Fish and Wildlife Research Unit, University of Washington, Seattle.
- Jones, E. L. 2004. Dietary Evenness, Prey Choice, and Human–Environment Interactions. *Journal of Archaeological Science* 31:307–317.
- Kadane, J. B. 1988. Possible Statistical Contributions to Paleoethnobotany. In *Current Paleoethnobotany*, edited by C. A. Hastorf and V. S. Popper, pp. 206–214. University of Chicago Press, Chicago.
- Kehoe, T. F., and A. B. Kehoe. 1960. Observations on the Butchering Technique at a Prehistoric Bison-Kill in Montana. *American Antiquity* 25:421–423.
- Kent, S. 1981. The Dog: An Archaeologist's Best Friend or Worst Enemy – The Spatial Distribution of Faunal Remains. *Journal of Field Archaeology* 8:367–372.
- Kerrich, J. E., and D. L. Clarke. 1967. Notes on the Possible Misuse and Errors of Cumulative Percentage Frequency Graphs for the Comparison of Prehistoric Artefact Assemblages. *Proceedings of the Prehistoric Society* 33:57–69.
- Kidwell, S. M. 2001. Preservation of Species Abundance in Marine Death Assemblages. *Science* 294:1091–1094.
- Kidwell, S. M. 2002. Time-Averaged Molluscan Death Assemblages: Palimpsests of Richness, Snapshots of Abundance. *Geology* 30:803–806.
- Kintigh, K. W. 1984. Measuring Archaeological Diversity by Comparison with Simulated Assemblages. *American Antiquity* 49:44–54.
- Klein, R. G. 1978. The Fauna and Overall Interpretation of the “Cutting 10” Acheulian Site at Elandsfontein (Hopefield), Southwestern Cape Province, South Africa. *Quaternary Research* 10:69–83.
- Klein, R. G. 1980. The Interpretation of Mammalian Faunas from Stone-Age Archeological Sites, with Special Reference to Sites in the Southern Cape Province, South Africa. In *Fossils in the Making*, edited by A. K. Behrensmeyer and A. P. Hill, pp. 223–246. University of Chicago Press, Chicago.

- Klein, R. G. 1989. Why Does Skeletal Part Representation Differ Between Smaller and Larger Bovids at Klasies River Mouth and Other Archeological Sites? *Journal of Archaeological Science* 16:363–381.
- Klein, R. G., and K. Cruz-Uribe. 1984. *The Analysis of Animal Bones from Archeological Sites*. University of Chicago Press, Chicago.
- Klein, R. G., and K. Cruz-Uribe. 1991. The Bovids from Elandsfontein, South Africa, and Their Implications for the Age, Palaeoenvironment, and Origins of the Site. *African Archaeological Review* 9:21–79.
- Klein, R. G., and K. Cruz-Uribe. 1994. The Paleolithic Mammalian Fauna from the 1910–14 Excavations at El Castillo Cave (Cantabria). *Museo y Centro de Investigación de Altamira, Monografías* 17:141–158.
- Klippel, W. E., L. M. Snyder, and P. W. Parmalee. 1987. Taphonomy and Archaeologically Recovered Mammal Bone from Southeast Missouri. *Journal of Ethnobiology* 7:155–169.
- Koch, C. F. 1987. Prediction of Sample Size Effects on the Measured Temporal and Geographic Distribution Patterns of Species. *Paleobiology* 13:100–107.
- Kooymann, B. 2004. Identification of Marrow Extraction in Zooarchaeological Assemblages Based on Fracture Patterns. In *Archaeology on the Edge: New Perspectives from the Northern Plains*, edited by B. Kooyman and J. H. Kelley, pp. 187–209. Canadian Archaeological Association Occasional Paper No. 4. Calgary, Alberta.
- Korth, W. W. 1979. Taphonomy of Microvertebrate Fossil Assemblages. *Annals of the Carnegie Museum* 48:235–285.
- Kowalewski, M. 2002. The Fossil Record of Predation: An Overview of Analytical Methods. In *The Fossil Record of Predation*, edited by M. Kowalewski and P. H. Kelley, pp. 3–42. Paleontological Society Paper No. 8. New Haven, Connecticut.
- Kowalewski, M., M. Carroll, L. Casazza, N. S. Gupta, B. Hannisdal, A. Hendy, R. A. Krause, Jr., M. LaBarbera, D. G. Lazo, C. Messina, S. Puchalski, T. A. Rothfus, J. Sälgeback, J. Stempien, R. C. Terry, and A. Tomasovich. 2003. Quantitative Fidelity of Brachiopod–Mollusk Assemblages from Modern Subtidal Environments of San Juan Islands, USA. *Journal of Taphonomy* 1:43–66.
- Kowalewski, M., G. A. Goodfriend, and K. W. Flessa. 1998. High-Resolution Estimates of Temporal Mixing Within Shell Beds: The Evils and Virtues of Time-Averaging. *Paleobiology* 24:287–304.
- Kowalewski, M., and A. P. Hoffmeister. 2003. Sieves and Fossils: Effects of Mesh Size on Paleontological Patterns. *Palaios* 18:460–469.
- Krantz, G. S. 1968. A New Method of Counting Mammal Bones. *American Journal of Archaeology* 72:286–288.
- Kranz, P. M. 1977. A Model for Estimating Standing Crop in Ancient Communities. *Paleobiology* 3:415–421.
- Krumbein, W. C. 1965. Sampling in Paleontology. In *Handbook of Paleontological Techniques*, edited by B. Kummel and D. Raup, pp. 137–150. W. H. Freeman & Co., San Francisco.
- Kuehne, W. G. 1971. Collecting Vertebrate Fossils by the Henkel Process. *Curator* 14:175–179.

- Kusmer, K. D. 1990. Taphonomy of Owl Pellet Deposition. *Journal of Paleontology* 64:629–637.
- Lande, R. 1996. Statistics and Partitioning of Species Diversity, and Similarity Among Multiple Communities. *Oikos* 76:5–13.
- Landon, D. B. 1996. Feeding Colonial Boston: A Zooarchaeological Study. *Historical Archaeology* 30:1–153.
- Lapham, H. A. 2005. *Hunting for Hides: Deerskins, Status, and Cultural Change in the Protohistoric Appalachians*. University of Alabama Press, Tuscaloosa.
- Lawrence, B. 1973. Problems in the Inter-Site Comparison of Faunal Remains. In *Domestikationsforschung und Geschichte der Haustiere*, edited by J. Matolcsi, pp. 397–402. Akademiai Kiado, Budapest.
- Lawson, J. D. 1999. Autecology and Communities. In *Paleocommunities: A Case Study from the Silurian and Lower Devonian*, edited by A. J. Boucot and J. D. Lawson, pp. 7–12. Cambridge University Press, Cambridge, UK.
- Leonard, R. D. 1987. Incremental Sampling in Artifact Analysis. *Journal of Field Archaeology* 14:498–500.
- Leonard, R. D. 1989. *Anasazi Faunal Exploitation: Prehistoric Subsistence on Northern Black Mesa, Arizona*. Center for Archaeological Investigations Occasional Paper No. 13. Southern Illinois University, Carbondale.
- Leonard, R. D. 1997. The Sample Size–Richness Relation: A Comment on Plog and Hegmon. *American Antiquity* 62:713–716.
- Leonardi, G., and G. L. Dell'Arte. 2006. Food Habits of the Barn Owl (*Tyto alba*) in a Steppe Area of Tunisia. *Journal of Arid Environments* 65:677–681.
- Lepofsky, D., and K. Lertzman. 2005. More on Sampling for Richness and Diversity in Archaeobiological Assemblages. *Journal of Ethnobiology* 25:175–188.
- Lie, R. W. 1980. Minimum Number of Individuals from Osteological Samples. *Norwegian Archaeological Review* 13:24–30.
- Lie, R. W. 1983. Reply (to Wild and Nichol). *Norwegian Archaeological Review* 16:49.
- Livingston, S. D. 1984. Faunal Analysis. In *Archaeological Investigations at Sites 45-OK-2 and 45-OK-2A, Chief Joseph Dam Project, Washington*, edited by S. K. Campbell, pp. 191–205. Office of Public Archaeology, Institute for Environmental Studies, University of Washington, Seattle. Report to the U.S. Army Corps of Engineers, Seattle District.
- Loreau, M. 2000. Are Communities Saturated? On the Relationship Between  $\alpha$ ,  $\beta$  and  $\gamma$  Diversity. *Ecology Letters* 3:73–76.
- Lorrain, D. 1968. Analysis of the Bison Bones from Bonfire Shelter. In *Bonfire Shelter: A Stratified Bison Kill Site, Val Verde County, Texas*, edited by D. S. Dibble and D. Lorrain, pp. 78–132. Texas Memorial Museum, Miscellaneous Papers No. 1.
- Lubinski, P. M. 2000. Of Bison and Lesser Mammals: Prehistoric Hunting Patterns in the Wyoming Basin. In *Intermountain Archaeology*, edited by D. B. Madsen and M. D. Metcalf, pp. 176–188. University of Utah Anthropological Papers No. 122. Salt Lake City.
- Lundelius, E., Jr. 1964. The Use of Vertebrates in Paleoecological Reconstructions. *Fort Burgwin Research Center Publication* 3:26–31.

- Lupo, K. D., and J. F. O'Connell. 2002. Cut and Tooth Mark Distributions on Large Animal Bones: Ethnoarchaeological Data from the Hadza and Their Implications for Current Ideas About Early Human Carnivory. *Journal of Archaeological Science* 29:85–109.
- Lyman, R. L. 1977. Analysis of Historic Faunal Remains. *Historical Archaeology* 11:67–73.
- Lyman, R. L. 1979. Available Meat from Faunal Remains: A Consideration of Techniques. *American Antiquity* 44:536–546.
- Lyman, R. L. 1984. Bone Density and Differential Survivorship of Fossil Classes. *Journal of Anthropological Archaeology* 3:259–299.
- Lyman, R. L. 1987a. Archaeofaunas and Butchery Studies: A Taphonomic Perspective. *Advances in Archaeological Method and Theory* 10:249–337.
- Lyman, R. L. 1987b. On Zooarchaeological Measures of Socioeconomic Position and Cost-Efficient Meat Purchases. *Historical Archaeology* 21:58–66.
- Lyman, R. L. 1988. Zooarchaeology of 45 DO189. In *Archaeological Investigations at River Mile 590: The Excavations at 45 DO189*, edited by J. R. Galm and R. L. Lyman, pp. 97–141. Eastern Washington University Reports in Archaeology and History 100–61. Cheney.
- Lyman, R. L. 1989. Taphonomy of Cervids Killed by the 18 May 1980 Volcanic Eruption of Mount St. Helens, Washington, U.S.A. In *Bone Modification*, edited by R. Bonnichsen and M. Sorg, pp. 149–167. University of Maine Center for the Study of Early Man, Orono.
- Lyman, R. L. 1991. *Prehistory of the Oregon Coast: The Effects of Excavation Strategies and Assemblage Size on Archaeological Inquiry*. Academic Press, San Diego.
- Lyman, R. L. 1992a. Review of “The Economic Prehistory of Namu” by Aubrey Cannon. *Canadian Journal of Archaeology* 16:134–136.
- Lyman, R. L. 1992b. Prehistoric Seal and Sea-Lion Butchering on the Southern Northwest Coast. *American Antiquity* 57:246–261.
- Lyman, R. L. 1994a. Quantitative Units and Terminology in Zooarchaeology. *American Antiquity* 59:36–71.
- Lyman, R. L. 1994b. Relative Abundances of Skeletal Specimens and Taphonomic Analysis of Vertebrate Remains. *Palaios* 9:288–298.
- Lyman, R. L. 1994c. *Vertebrate Taphonomy*. Cambridge University Press, Cambridge, UK.
- Lyman, R. L. 1995a. Determining When Rare (Zoo)Archaeological Phenomena Are Truly Absent. *Journal of Archaeological Method and Theory* 2:369–424.
- Lyman, R. L. 1995b. A Study of Variation in the Prehistoric Butchery of Large Artiodactyls. In *Ancient Peoples and Landscapes*, edited by E. Johnson, pp. 233–253. Museum of Texas Tech University, Lubbock.
- Lyman, R. L. 2003a. Pinniped Behavior, Foraging Theory, and the Depression of Metapopulations and Nondepression of a Local Population on the Southern Northwest Coast of North America. *Journal of Anthropological Archaeology* 22:376–388.
- Lyman, R. L. 2003b. The Influence of Time Averaging and Space Averaging on the Application of Foraging Theory in Zooarchaeology. *Journal of Archaeological Science* 30:595–610.
- Lyman, R. L. 2004a. Prehistoric Biogeography, Abundance, and Phenotypic Plasticity of Elk (*Cervus elaphus*) in Washington State. In *Zooarchaeology and Conservation Biology*, edited by R. L. Lyman and K. P. Cannon, pp. 136–163. University of Utah Press, Salt Lake City.

- Lyman, R. L. 2004b. Late-Quaternary Diminution and Abundance of Prehistoric Bison (*Bison* sp.) in Eastern Washington State, U.S.A. *Quaternary Research* 62:76–85.
- Lyman, R. L. 2004c. The Concept of Equifinality in Taphonomy. *Journal of Taphonomy* 2:15–26.
- Lyman, R. L. 2004d. Aboriginal Overkill in the Intermountain West of North America: Zooarchaeological Tests and Implications. *Human Nature* 15:169–208.
- Lyman, R. L. 2005a. Zooarchaeology. In *Handbook of Archaeological Methods*, edited by C. Chippindale and H. D. G. Maschner, pp. 835–870. Altimira Press, Walnut Creek, California.
- Lyman, R. L. 2005b. Analyzing Cut Marks: Lessons from Artiodactyl Remains in the Northwestern United States. *Journal of Archaeological Science* 32:1722–1732.
- Lyman, R. L. 2006a. Identifying Bilateral Pairs of Deer (*Odocoileus* sp.) Bones: How Symmetrical is Symmetrical Enough? *Journal of Archaeological Science* 33:1256–1265.
- Lyman, R. L. 2006b. Late Prehistoric and Early Historic Abundance of Columbian White-Tailed Deer, Portland Basin, Washington and Oregon, U.S.A. *Journal of Wildlife Management* 70:278–282.
- Lyman, R. L. 2006c. Archaeological Evidence of Anthropogenically Induced Twentieth-Century Diminution of North American Wapiti (*Cervus elaphus*). *American Midland Naturalist* 156:88–98.
- Lyman, R. L., and K. A. Ames. 2004. Sampling to Redundancy in Zooarchaeology: Lessons from the Portland Basin, Northwestern Oregon and Southwestern Washington. *Journal of Ethnobiology* 24:329–346.
- Lyman, R. L., and K. A. Ames. 2007. On the Use of Species-Area Curves to Detect the Effects of Sample Size. *Journal of Archaeological Science* 34: 1985–1990.
- Lyman, R. L., and G. L. Fox. 1989. A Critical Evaluation of Bone Weathering as an Indication of Bone Assemblage Formation. *Journal of Archaeological Science* 16:293–317.
- Lyman, R. L., J. L. Harpole, C. Darwent, and R. Church. 2002. Prehistoric Occurrence of Pinnipeds in the Lower Columbia River. *Northwestern Naturalist* 83:1–6.
- Lyman, R. L., and R. J. Lyman. 2003. Lessons from Temporal Variation in the Mammalian Faunas from Two Collections of Owl Pellets in Columbia County, Washington. *International Journal of Osteoarchaeology* 13:150–156.
- Lyman, R. L., and M. J. O'Brien. 1987. Plow-Zone Zooarchaeology: Fragmentation and Identifiability. *Journal of Field Archaeology* 14:493–498.
- Lyman, R. L., and M. J. O'Brien. 1999. Americanist Stratigraphic Excavation and the Measurement of Culture Change. *Journal of Archaeological Method and Theory* 6:55–108.
- Lyman, R. L., and M. J. O'Brien. 2005. Within-Taxon Morphological Diversity as a Paleoenvironmental Indicator: Late-Quaternary *Neotoma* in the Bonneville Basin, Northwestern Utah. *Quaternary Research* 63:274–282.
- Lyman, R. L., E. Power, and R. J. Lyman. 2001. Ontogeny of Deer Mice (*Peromyscus maniculatus*) and Montane Voles (*Microtus montanus*) as Owl Prey. *American Midland Naturalist* 146:72–79.
- Lyman, R. L., E. Power, and R. J. Lyman. 2003. Quantification and Sampling of Faunal Remains in Owl Pellets. *Journal of Taphonomy* 1:3–14.

- Lyman, R. L., and J. Zehr. 2003. Archaeological Evidence of Mountain Beaver (*Aplodontia rufa*) Mandibles as Chisels and Engravers on the Northwest Coast. *Journal of Northwest Anthropology* 37:89–100.
- MacKenzie, D. I. 2005. What are the Issues with Presence–Absence Data for Wildlife Managers? *Journal of Wildlife Management* 69:849–860.
- Magurran, A. E. 1988. *Ecological Diversity and Its Measurement*. Princeton University Press, Princeton.
- Maltby, J. M. 1985. Assessing Variation in Iron Age and Roman Butchery Practices: The Need for Quantification. In *Paleobiological Investigations: Research Design, Methods and Data Analysis*, edited by N. R. J. Fieller, D. D. Gilbertson, and N. G. A. Ralph, pp. 19–30. British Archaeological Reports International Series 266. Oxford.
- Marean, C. W. 1992. Hunter to Herder: Large Mammal Remains from the Hunter-Gatherer Occupation at Enkapune Ya Muto Rockshelter, Central Rift, Kenya. *The African Archaeological Review* 10:65–127.
- Marean, C. W. 1995. Of Taphonomy and Zooarcheology. *Evolutionary Anthropology* 4:64–72.
- Marean, C. W., Y. Abe, C. J. Frey, and R. C. Randall. 2000. Zooarchaeological and Taphonomic Analysis of the Die Kelders Cave 1 Layers 10 and 11 Middle Stone Age Larger Mammal Fauna. *Journal of Human Evolution* 38:197–233.
- Marean, C. W., Y. Abe, P. Nilssen, and E. Stone. 2001. Estimating the Minimum Number of Skeletal Elements (MNE) in Zooarchaeology: A Review and A New Image-Analysis GIS Approach. *American Antiquity* 66:333–348.
- Marean, C. W., M. Domínguez-Rodrigo, and T. R. Pickering. 2004. Skeletal Element Equifinality in Zooarchaeology Begins with Method: The Evolution and Current Status of the “Shaft Critique.” *Journal of Taphonomy* 2:69–98.
- Marean, C. W., and C. L. Ehrhardt. 1995. Paleoanthropological and Paleoecological Implications of the Taphonomy of a Sabertooth’s Den. *Journal of Human Evolution* 29:515–547.
- Marean, C. W., and C. J. Frey. 1997. Animal Bones from Caves to Cities: Reverse Utility Curves as Methodological Artifacts. *American Antiquity* 62:698–711.
- Marean, C. W., and S. Y. Kim. 1998. Mousterian Large-Mammal Remains from Kobeh Cave: Behavioral Implications for Neanderthals and Early Modern Humans. *Current Anthropology* 39:S79–S113.
- Marean, C. W., and L. M. Spencer. 1991. Impact of Carnivore Ravaging on Zooarchaeological Measures of Element Abundance. *American Antiquity* 56:645–658.
- Marshall, F., and T. Pilgram. 1991. Meat versus Within-Bone Nutrients: Another Look at the Meaning of Body-Part Representation in Archaeological Sites. *Journal of Archaeological Science* 18:149–163.
- Marshall, F., and T. Pilgram. 1993. NISP vs. MNI in Quantification of Body-Part Representation. *American Antiquity* 58:261–269.
- Marti, C. D. 1987. Raptor Food Habits Studies. In *Raptor Management Techniques Manual*, edited by B. A. G. Pendleton, B. A. Millsap, K. W. Cline, and D. M. Bird, pp. 67–80. National Wildlife Federation, Washington, DC.

- Matthews, T. 2002. South African Micromammals and Predators: Some Comparative Results. *Archaeometry* 44:363–370.
- Mayhew, D. F. 1977. Avian Predators as Accumulators of Fossil Mammal Material. *Boreas* 6:25–31.
- McCartney, P. H., and M. F. Glass. 1990. Simulation Models and the Interpretation of Archaeological Diversity. *American Antiquity* 55:521–536.
- McClure, S. B. 2004. Small Mammal Procurement in Coastal Contexts: A California Perspective. *Journal of California and Great Basin Anthropology* 24:207–232.
- McKenna, M. C. 1962. Collecting Small Fossils by Washing and Screening. *Curator* 5:221–235.
- McMahon, T. A. 1975. Allometry and Biomechanics: Limb Bones in Adult Ungulates. *American Naturalist* 109:547–563.
- McMeekan, C. P. 1940. Growth and Development in the Pig, with Special Reference to Carcass Quality Characters, Part I. *Journal of Agricultural Science* 30:276–343.
- Mendoza, M., C. M. Janis, and P. Palmqvist. 2006. Estimating the Body Mass of Extinct Ungulates: A Study on the Use of Multiple Regression. *Journal of Zoology* 270:90–101.
- Miller, A. I., and M. Foote. 1996. Calibrating the Ordovician Radiation of Marine Life: Implications for Phanerozoic Diversity Trends. *Paleobiology* 22:304–309.
- Miller, G. R., and A. L. Gill. 1990. Zooarchaeology at Pirincay, a Formative Period Site in Highland Ecuador. *Journal of Field Archaeology* 17:49–68.
- Milo, R. G. 1998. Evidence for Hominid Predation at Klasies River Mouth, South Africa, and Its Implications for the Behaviour of Early Modern Humans. *Journal of Archaeological Science* 25:99–133.
- Mollhagen, T. R., R. W. Wiley, and R. L. Packard. 1972. Prey Remains in Golden Eagle Nests: Texas and New Mexico. *Journal of Wildlife Management* 36:784–792.
- Monahan, C. M. 1999. Comparing Apples and Oranges in the Plio-Pleistocene: Methodological Comments on “Meat-Eating by Early Hominids at the FLK 22 *Zinjanthropus* Site, Olduvai Gorge (Tanzania): An Experimental Approach Using Cut-Mark Data.” *Journal of Human Evolution* 37:789–792.
- Monks, G. G. 2000. How Much is Enough? An Approach to Sampling Ichthyofaunas. *Ontario Archaeology* 69:65–75.
- Moore, P. D., J. A. Webb, and M. E. Collinson. 1991. *Pollen Analysis*, 2nd ed. Blackwell Scientific, Oxford.
- Morlan, R. E. 1983. Counts and Estimates of Taxonomic Abundance in Faunal Remains: Microtine Rodents from Bluefish Cave I. *Canadian Journal of Archaeology* 7:61–76.
- Morlan, R. E. 1994. Bison Bone Fragmentation and Survivorship: A Comparative Method. *Journal of Archaeological Science* 21:797–807.
- Morrison, D. A. 1997. *Caribou Hunters in the Western Arctic: Zooarchaeology of the Rita-Claire and Bison Skull Site*. Archaeological Survey of Canada Mercury Series Paper No. 157. Canadian Museum of Civilization, Hull, Quebec.
- Muir, R. J., and J. C. Driver. 2002. Scale of Analysis and Zooarchaeological Interpretation: Pueblo III Faunal Variation in the Northern San Juan Region. *Journal of Anthropological Archaeology* 21:165–199.

- Munro, N. D., and G. Bar-Oz. 2005. Gazelle Bone Fat Processing in the Levantine Epipaleolithic. *Journal of Archaeological Science* 32:223–239.
- Nagaoka, L. 2001. Using Diversity Indices to Measure Changes in Prey Choice at the Shag River Mouth Site, Southern New Zealand. *International Journal of Osteoarchaeology* 11:101–111.
- Nagaoka, L. 2002. Explaining Subsistence Change in Southern New Zealand Using Foraging Theory Models. *World Archaeology* 34:84–102.
- Nagaoka, L. 2005a. Declining Foraging Efficiency and Moa Carcass Exploitation in Southern New Zealand. *Journal of Archaeological Science* 32:1328–1338.
- Nagaoka, L. 2005b. Differential Recovery of Pacific Island Fish Remains. *Journal of Archaeological Science* 32:941–955.
- Nance, J. D. 1983. Regional Sampling in Archaeological Survey: The Statistical Perspective. *Advances in Archaeological Method and Theory* 6:289–356.
- Needs-Howarth, S. 1995. Quantifying Animal Food Diet: A Comparison of Four Approaches Using Bones from a Prehistoric Iroquoian Village. *Ontario Archaeology* 60:92–101.
- Nichol, R. K., and G. A. Creak. 1979. Matching Paired Elements Among Archaeological Bone Remains: A Computer Procedure and Some Practical Limitations. *Newsletter of Computer Archaeology* 14:6–17.
- Nichol, R. K., and C. J. Wild. 1984. “Numbers of Individuals” in Faunal Analysis: The Decay of Fish Bone in Archaeological Sites. *Journal of Archaeological Science* 11:35–51.
- Nichols, J. D. 1992. Capture–Recapture Models: Using Marked Animals to Study Population Dynamics. *BioScience* 42:94–102.
- Noe-Nygaard, N. 1977. Butchering and Marrow Fracturing as a Taphonomic Factor in Archaeological Deposits. *Paleobiology* 3:218–237.
- Noe-Nygaard, N. 1989. Man-Made Trace Fossils on Bones. *Human Evolution* 4:461–491.
- Noodle, B. 1973. Determination of the Body Weight of Cattle from Bone Measurements. In *Domestikationsforschung und Geschichte der Haustiere*, edited by J. Matolcsi, pp. 377–389. Akadémiai Kiadó, Budapest.
- O’Connell, J. F. 1987. Alyawara Site Structure and Its Archaeological Implications. *American Antiquity* 52:74–108.
- O’Connell, J. F., K. Hawkes, K. D. Lupo, and N. G. Blurton Jones. 2003. Another Reply to Domínguez-Rodrigo. *Journal of Human Evolution* 45:417–419.
- O’Connell, J. F., and K. D. Lupo. 2003. Reply to Domínguez-Rodrigo. *Journal of Archaeological Science* 30:387–390.
- O’Connor, T. 2000. *The Archaeology of Animal Bones*. Texas A&M University Press, College Station.
- O’Connor, T. P. 2001. Animal Bone Quantification. In *Handbook of Archaeological Sciences*, edited by D. R. Brothwell and A. M. Pollard, pp. 703–710. John Wiley and Sons, Chichester.
- O’Connor, T. P. 2003. *The Analysis of Urban Animal Bone Assemblages: A Handbook for Archaeologists*. Archaeology of York 19(2):69–224. Council for British Archaeology, York.
- Odum, E. P. 1971. *Fundamentals of Ecology*, 3rd ed. W. B. Saunders, Philadelphia.

- Oliver, J. S. 1994. Estimates of Hominid and Carnivore Involvement in the FLK Zinjanthropus Fossil Assemblage: Some Socioecological Implications. *Journal of Human Evolution* 27:267–294.
- Olsen, S. L., and P. Shipman. 1988. Surface Modification on Bone: Trampling versus Butchery. *Journal of Archaeological Science* 15:535–553.
- Olszewski, T. 1999. Taking Advantage of Time Averaging. *Paleobiology* 25:226–238.
- Olszewski, T. D. 2004. A Unified Mathematical Framework for the Measurement of Richness and Evenness Within and Among Multiple Communities. *Oikos* 104:377–387.
- Olszewski, T. D., and S. M. Kidwell. 2007. The Preservational Fidelity of Evenness in Molluscan Death Assemblages. *Paleobiology* 33:1–23.
- Orchard, T. J. 2005. The Use of Statistical Size Estimations in Minimum Number Calculations. *International Journal of Osteoarchaeology* 15:351–359.
- Orton, C. 2000. *Sampling in Archaeology*. Cambridge University Press, Cambridge, UK.
- Osman, R. W., and R. B. Whitlatch. 1978. Patterns of Species Diversity: Fact or Artifact? *Paleobiology* 4:41–54.
- Palmer, A. R. 1986. Inferring Relative Levels of Genetic Variability in Fossils: The Link Between Heterozygosity and Fluctuating Asymmetry. *Paleobiology* 12:1–5.
- Palmer, A. R. 1994. Fluctuating Asymmetry Analyses: A Primer. In *Developmental Instability*, edited by T. A. Markow, pp. 335–364. Kluwer, Dordrecht.
- Palmer, A. R. 1996. Waltzing with Asymmetry. *BioScience* 46:518–532.
- Palmer, A. R., and C. Strobeck. 1986. Fluctuating Asymmetry: Measurement, Analysis, Patterns. *Annual Review of Ecology and Systematics* 17:391–421.
- Palmer, M. W. 1990. The Estimation of Species Richness by Extrapolation. *Ecology* 71:1195–1198.
- Pankakoski, E. 1985. Epigenetic Asymmetry as an Ecological Indicator in Muskrats. *Journal of Mammalogy* 66:52–57.
- Partlow, M. A. 2006. Sampling Fish Bones: A Consideration of the Importance of Screen Size and Disposal Context in the North Pacific. *Arctic Anthropology* 43:67–79.
- Patterson, B. D. 1987. The Principle of Nested Subsets and Its Implications for Biological Conservation. *Conservation Biology* 1:323–334.
- Patterson, B. D., and W. Atmar. 1986. Nested Subsets and the Structure of Insular Mammalian Faunas and Archipelagos. *Biological Journal of the Linnean Society* 28:65–82.
- Pavao-Zuckerman, B. 2007. Deerskins and Domesticates: Creek Subsistence and Economic Strategies in the Historic Period. *American Antiquity* 72:5–33.
- Payne, S. 1972. Partial Recovery and Sample Bias: The Results of Some Sieving Experiments. In *Papers in Economic Prehistory*, edited by E. S. Higgs, pp. 49–64. Cambridge University Press, Cambridge, UK.
- Payne, S. 1975. Partial Recovery and Sample Bias. In *Archaeozoological Studies*, edited by A. T. Clason, pp. 7–17. North Holland, Amsterdam.
- Pearson, O. P., and A. K. Pearson. 1947. Owl Predation in Pennsylvania, with Notes on the Small Mammals of Delaware County. *Journal of Mammalogy* 28:137–147.

- Peet, R. K. 1974. The Measurement of Species Diversity. *Annual Review of Ecology and Systematics* 5:285–307.
- Perkins, D., Jr. 1973. A Critique on the Methods of Quantifying Faunal Remains from Archaeological Sites. In *Domestikationsforschung und Geschichte der Haustiere*, edited by J. Matolcsi, pp. 367–369. Akademiai Kiado, Budapest.
- Peterson, C. H. 1977. The Paleoecological Significance of Undetected Short-Term Temporal Variability. *Journal of Paleontology* 51:976–981.
- Pettigrew, R. M. 1981. *A Prehistoric Cultural Sequence in the Portland Basin of the Lower Columbia Valley*. University of Oregon Anthropological Papers No. 22. Eugene.
- Pianka, E. R. 1978. *Evolutionary Ecology*, 2nd ed. Harper and Row, New York.
- Pickering, T. R., and C. P. Egeland. 2006. Experimental Patterns of Hammerstone Percussion Damage on Bones: Implications for Inferences of Carcass Processing by Humans. *Journal of Archaeological Science* 33:459–469.
- Pickering, T. R., C. W. Marean, and M. Domínguez-Rodrigo. 2003. Importance of Limb Bone Shaft Fragments in Zooarchaeology: A Response to “On *in situ* Attrition and Vertebrate Body Part Profiles” (2002), by M. C. Stiner. *Journal of Archaeological Science* 30:1469–1482.
- Pilgram, T., and F. Marshall. 1995. Bone Counts and Statisticians: A Reply to Ringrose. *Journal of Archaeological Science* 22:93–97.
- Pimm, S. L., and J. H. Lawton. 1998. Planning for Biodiversity. *Science* 279:2068–2069.
- Plug, C., and I. Plug. 1990. MNI Counts as Estimates of Species Abundance. *South African Archaeological Bulletin* 45:53–57.
- Plug, I. 2004. Resource Exploitation: Animal Use During the Middle Stone Age at Sibudu Cave, KwaZulu-Natal. *South African Journal of Science* 100:151–158.
- Plug, I., and S. Badenhorst. 2006. Notes on the Fauna from Three Late Iron Age Mega-Sites, Boitsemagano, Molokwane, and Mabjanamatshwana, North West Province, South Africa. *South African Archaeological Bulletin* 61:57–67.
- Plug, I., and C. G. Sampson. 1996. European and Bushman Impacts on Karoo Fauna in the Nineteenth Century: An Archaeological Perspective. *South African Archaeological Bulletin* 51:26–31.
- Pobiner, B. L., and D. R. Braun. 2005. Strengthening the Inferential Link Between Cutmark Frequency Data and Oldowan Hominid Behavior: Results from Modern Butchery Experiments. *Journal of Taphonomy* 3:107–119.
- Popper, V. S. 1988. Selecting Quantitative Measures in Paleoethnobotany. In *Current Paleoethnobotany: Analytical Methods and Cultural Interpretations of Archaeological Plant Remains*, edited by C. A. Hastorf and V. S. Popper, pp. 53–71. University of Chicago Press, Chicago, Illinois.
- Potter, S. L. 2005. The Physics of Cutmarks. *Journal of Taphonomy* 3:91–106.
- Potts, R. 1986. Temporal Span of Bone Accumulations at Olduvai Gorge and Implications for Early Hominid Foraging Behavior. *Paleobiology* 12:25–31.
- Potts, R. B. 1988. *Early Hominid Activities at Olduvai*. Aldine de Gruyter, New York.
- Pozorski, S. 1979. Late Prehistoric Llama Remains from the Moche Valley, Peru. *Annals of the Carnegie Museum* 48:139–170.

- Prange, H. D., J. F. Anderson, and H. Rahn. 1979. Scaling of Skeletal Body Mass in Birds and Mammals. *American Naturalist* 113:103–122.
- Prummel, W. 2003. Animal Remains from the Hellenistic Town of New Halos in the Almirós Plain, Thessaly. In *Zooarchaeology in Greece: Recent Advances*, edited by E. Kotjabopoulou, Y. Hamilakis, P. Halstead, C. Gamble, and P. Elefanti, pp. 153–159. British School at Athens Studies 9, London.
- Purdue, J. R. 1980. Clinal Variation of Some Mammals During the Holocene in Missouri. *Quaternary Research* 13:242–258.
- Purdue, J. R. 1982. Methods of Determining Sex and Body Size in Prehistoric Samples of White-tailed Deer (*Odocoileus virginianus*). *Transactions of the Illinois State Academy of Science* 76:351–357.
- Purdue, J. R. 1986. The Size of White-Tailed Deer (*Odocoileus virginianus*) During the Archaic Period in Central Illinois. In *Foraging, Collecting, and Harvesting: Archaic Period Subsistence and Settlement in the Eastern Woodlands*, edited by S. W. Neusius, pp. 65–95. Center for Archaeological Investigations Occasional Paper No. 6. Southern Illinois University, Carbondale.
- Purdue, J. R. 1987. Estimation of Body Weight of White-tailed Deer (*Odocoileus virginianus*) from Bone Size. *Journal of Ethnobiology* 7:1–12.
- Purdue, J. R., B. W. Styles, and M. C. Masulis. 1989. Faunal Remains and White-Tailed Deer Exploitation from a Late Woodland Upland Encampment: The Boschert Site (23SC609), St. Charles County, Missouri. *Midcontinental Journal of Archaeology* 14:146–163.
- Quimby, G. I. 1960. Habitat, Culture, and Archaeology. In *Essays in the Science of Culture in Honor of Leslie A. White*, edited by G. E. Cole and R. L. Carneiro, pp. 380–389. Thomas Y. Crowell Company, New York.
- Quitmyer, I. R., and E. J. Reitz. 2006. Marine Trophic Levels Targeted Between AD 300 and 1500 on the Georgia Coast, USA. *Journal of Archaeological Science* 33:806–822.
- Rackham, J. 1994. *Animal Bones*. University of California Press, Berkeley.
- Rapson, D. J., and L. C. Todd. 1992. Conjoins, Contemporaneity, and Site Structure: Distributional Analyses of the Bugas-Holding Site. In *Piecing Together the Past: Applications of Refitting Studies in Archaeology*, edited by J. L. Hofman and J. G. Enloe, pp. 238–263. British Archaeological Reports International Series 578. Oxford.
- Raup, D. M. 1972. Taxonomic Diversity During the Phanerozoic. *Science* 177:1065–1071.
- Raup, D. M., and R. E. Crick. 1979. Measurement of Faunal Similarity in Paleontology. *Journal of Paleontology* 53:1213–1227.
- Reed, C. A. 1963. Osteo-Archaeology. In *Science in Archaeology*, edited by D. R. Brothwell and E. S. Higgs, pp. 204–216. Thames and Hudson, London.
- Reitz, E. J. 1988. Preceramic Animal Use on the Central Coast. In *Economic Prehistory of the Central Andes*, edited by E. S. Wing and J. C. Wheeler, pp. 31–55. British Archaeological Reports International Series 427. Oxford.
- Reitz, E. J. 1994. Zooarchaeological Analysis of a Free African Community: Gracia Real de Santa Teresa de Mose. *Historical Archaeology* 28:23–40.

- Reitz, E. J. 2003. Resource Use Through Time at Paloma, Peru. In *Zooarchaeology: Papers to Honor Elizabeth S. Wing*, edited by F. W. King and C. M. Porter, pp. 65–80. Bulletin of the Florida State Museum 44. Gainesville.
- Reitz, E. J., and D. Cordier. 1983. Use of Allometry in Zooarchaeological Analysis. In *Animals and Archaeology: 2. Shell Middens, Fishes and Birds*, edited by C. Grigson and J. Clutton-Brock, pp. 237–252. British Archaeological Reports International Series 183. Oxford.
- Reitz, E., and N. Honerkamp. 1983. British Colonial Subsistence Strategy on the Southeastern Coastal Plain. *Historical Archaeology* 17:4–26.
- Reitz, E. J., I. R. Quitmyer, H. S. Hale, S. J. Scudder, and E. S. Wing. 1987. Application of Allometry to Zooarchaeology. *American Antiquity* 52:304–317.
- Reitz, E. J., and E. S. Wing. 1999. *Zooarchaeology*. Cambridge University Press, Cambridge, UK.
- Reynolds, P. S. 2002. How Big is Giant? The Importance of Method in Estimating Body Size of Extinct Mammals. *Journal of Mammalogy* 83:321–332.
- Rhode, D. 1988. Measurement of Archaeological Diversity and Sample-Size Effect. *American Antiquity* 53:708–716.
- Ricklefs, R. E. 1979. *Ecology*, 2nd ed. Chiron Press, New York.
- Ringrose, T. J. 1993. Bone Counts and Statistics: A Critique. *Journal of Archaeological Science* 20:121–157.
- Ringrose, T. J. 1995. Response to Pilgram and Marshall “Bone Counts and Statisticians: A Reply to Ringrose.” *Journal of Archaeological Science* 22:99–102.
- Rogers, A. R. 2000a. Analysis of Bone Counts by Maximum Likelihood. *Journal of Archaeological Science* 27:111–125.
- Rogers, A. R. 2000b. On Equifinality in Faunal Analysis. *American Antiquity* 65:709–723.
- Rogers, A. R., and J. M. Broughton. 2001. Selective Transport of Animal Parts by Ancient Hunters: A New Statistical Method and an Application to the Emeryville Shellmound Fauna. *Journal of Archaeological Science* 28:763–773.
- Saleeby, B. 1983. *Prehistoric Settlement Patterns in the Portland Basin on the Lower Columbia River: Ethnohistoric, Archaeological and Biogeographic Perspectives*. Ph.D. dissertation, University of Oregon, Eugene.
- Sanders, H. L. 1968. Marine Benthic Diversity: A Comparative Study. *American Naturalist* 48:675–706.
- Scheiner, S. M. 2003. Six Types of Species–Area Curves. *Global Ecology and Biogeography* 12:441–447.
- Schindel, D. E. 1980. Microstratigraphic Sampling and the Limits of Paleontologic Resolution. *Paleobiology* 6:408–426.
- Schmitt, D. N., and K. D. Lupo. 1995. On Mammalian Taphonomy, Taxonomic Diversity, and Measuring Subsistence Data in Zooarchaeology. *American Antiquity* 60:496–514.
- Schoereder, J. H., C. Galbiati, C. R. Ribas, T. G. Sobrinho, C. F. Sperber, O. DeSouza, and C. Lopes-Andrade. 2004. Should We Use Proportional Sampling for Species–Area Studies? *Journal of Biogeography* 31:1219–1226.

- Schulz, P. D., and S. M. Gust. 1983. Faunal Remains and Social Status in 19th Century Sacramento. *Historical Archaeology* 17:44–53.
- Scott, E. M. 2003. Horticultural Hunters: Seasonally Abundant Animal Resources and Gender Roles in Late Prehistoric Iroquoian Subsistence Strategies. In *Zooarchaeology: Papers to Honor Elizabeth S. Wing*, edited by F. W. King and C. M. Porter, pp. 171–182. Bulletin of the Florida State Museum 44. Gainesville.
- Scott, K. M. 1982. Prediction of Body Weight of Fossil Artiodactyla. *Zoological Journal of the Linnaean Society* 77:199–215.
- Selvaggio, M. M. 1994. Carnivore Tooth Marks and Stone Tool Butchery Marks on Scavenged Bones: Archaeological Implications. *Journal of Human Evolution* 27:215–228.
- Selvaggio, M. M. 1998. Evidence for a Three-Stage Sequence of Hominid and Carnivore Involvement with Long Bones at FLK Zinjanthropus, Olduvai Gorge, Tanzania. *Journal of Archaeological Science* 25:191–202.
- Sepkoski, J. J., Jr. 1988. Alpha, Beta, or Gamma: Where Does All the Diversity Go? *Paleobiology* 14:221–234.
- Sepkoski, J. J., Jr. 1997. Biodiversity: Past, Present, and Future. *Journal of Paleontology* 71:533–539.
- Shaffer, B. S. 1992. Quarter Inch Screening: Understanding Biases in Recovery of Vertebrate Faunal Remains. *American Antiquity* 57:129–136.
- Shaffer, B. S., and B. W. Baker. 1999. Comments on James' Methodological Issues Concerning Analysis of Archaeofaunal Recovery and Screen Size Correction Factors. *Journal of Archaeological Science* 26:1181–1182.
- Sharp, N. D. 1990. Fremont and Anasazi Resource Selection: An Examination of Faunal Assemblage Variation in the Northern Southwest. *Kiva* 56:45–65.
- Shennan, S. 1988. *Quantifying Archaeology*. Academic Press, San Diego.
- Shipman, P., G. Foster, and M. Schoeninger. 1984. Burnt Bones and Teeth: An Experimental Study of Color, Morphology, Crystal Structure and Shrinkage. *Journal of Archaeological Science* 11:307–325.
- Shipman, P., and J. Rose. 1983. Early Hominid Hunting, Butchering, and Carcass Processing Behaviors: Approaches to the Fossil Record. *Journal of Anthropological Archaeology* 2:57–98.
- Shipman, P., and A. Walker. 1980. Bone-Collecting by Harvesting Ants. *Paleobiology* 6:496–502.
- Shotwell, J. A. 1955. An Approach to the Paleoecology of Mammals. *Ecology* 36:327–337.
- Shotwell, J. A. 1958. Inter-Community Relationships in Hemphillian (Mid-Pliocene) Mammals. *Ecology* 39:271–282.
- Simpson, G. G., A. Roe, and R. C. Lewontin. 1960. *Quantitative Zoology*, revised ed. Harcourt, Brace and Company, New York. (reprinted 2003 by Dover Publications, Mineola, New York)
- Smith, B., and J. B. Wilson. 1996. A Consumer's Guide to Evenness Indices. *Oikos* 76:70–82.
- Smith, B. D. 1975. Toward a More Accurate Estimation of Meat Yield of Animal Species at Archaeological Sites. In *Archaeozoological Studies*, edited by A. T. Clason, pp. 99–106. North-Holland, Amsterdam.
- Smith, E. P., P. M. Stewart, and J. Cairns, Jr. 1985. Similarities Between Rarefaction Methods. *Hydrobiologia* 120:167–170.

- Smith, R. J. 2002. Estimation of Body Mass in Paleontology. *Journal of Human Evolution* 43:271–287.
- Snyder, L. M., and W. E. Klippel. 2003. From Lerna to Kastro: Further Thoughts on Dogs as Food in Ancient Greece; Perceptions, Prejudices and Reinvestigations. In *Zooarchaeology in Greece: Recent Advances*, edited by E. Kotjabopoulou, Y. Hamilakis, P. Halstead, C. Gamble, and P. Elefanti, pp. 221–231. British School at Athens, London.
- Southwood, T. R. E. 1987. The Concept and Nature of the Community. In *Organization of Communities: Past and Present*, edited by J. H. R. Gee and P. S. Giller, pp. 3–27. Blackwell Scientific, Oxford.
- Sparks, B. W. 1961. The Ecological Interpretation of Quaternary Non-Marine Mollusca. *Proceedings of the Linnean Society of London* 172:71–80.
- Spellerberg, I. F., and P. J. Fedor. 2003. A Tribute to Claude Shannon (1916–2001) and a Plea for More Rigorous Use of Species Richness, Species Diversity and the ‘Shannon-Wiener’ Index. *Global Ecology and Biogeography* 12:177–179.
- Spencer, L. M., B. Van Valkenburgh, and J. M. Harris. 2003. Taphonomic Analysis of Large Mammals Recovered from the Pleistocene Rancho La Brea Tar Seeps. *Paleobiology* 29:561–575.
- Speth, J. D. 1983. *Bison Kills and Bone Counts: Decision Making by Ancient Hunters*. University of Chicago Press, Chicago.
- Staff, G., E. N. Powell, R. J. Stanton, Jr., and H. Cummins. 1985. Biomass: Is It a Useful Tool in Paleocommunity Reconstruction? *Lethaia* 18:209–232.
- Stahl, P. W. 2000. Archaeofaunal Accumulation, Fragmented Forests, and Anthropogenic Landscape Mosaics in the Tropical Lowlands of Prehispanic Ecuador. *Latin American Antiquity* 11:241–257.
- Stahl, P. W., and J. S. Athens. 2001. A High Elevation Zooarchaeological Assemblage from the Northern Andes of Ecuador. *Journal of Field Archaeology* 28:161–176.
- Steenhof, K. 1983. Prey Weights for Computing Percent Biomass in Raptor Diets. *Raptor Research* 17(1):15–27.
- Stephens, D. W., and J. R. Krebs. 1986. *Foraging Theory*. Princeton University Press, Princeton, New Jersey.
- Stevens, S. S. 1946. On the Theory of Scales of Measurement. *Science* 103:677–680.
- Stewart, F. L., and P. W. Stahl. 1977. Cautionary Note on Edible Meat Poundage Figures. *American Antiquity* 42:267–270.
- Stewart-Abernathy, L. C., and B. L. Ruff. 1989. A Good Man in Israel: Zooarchaeology and Assimilation in Antebellum Washington, Washington, Arkansas. *Historical Archaeology* 23:96–112.
- Stiner, M. C. 1991. Food Procurement and Transport by Human and Non-Human Predators. *Journal of Archaeological Science* 18:455–482.
- Stiner, M. C. 2002. On *in situ* Attrition and Vertebrate Body Part Profiles. *Journal of Archaeological Science* 29:979–991.
- Stiner, M. C. 2004. A Comparison of Photon Densitometry and Computed Tomography Parameters of Bone Density in Ungulate Body Part Profiles. *Journal of Taphonomy* 2:117–145.

- Stiner, M. C. 2005. *The Faunas of Hayonim Cave, Israel: A 200,000 Year Record of Paleolithic Diet, Demography, and Society*. American School of Prehistoric Research Bulletin 48. Peabody Museum of Archaeology and Ethnology, Harvard University, Cambridge, Massachusetts.
- Stiner, M. C., N. D. Munro, and T. A. Surovell. 2000. The Tortoise and the Hare: Small-Game Use, the Broad Spectrum Revolution, and Paleolithic Demography. *Current Anthropology* 41:39–73.
- Stiner, M. C., N. D. Munro, T. A. Surovell, E. Tchernov, and O. Bar-Yosef. 1999. Paleolithic Population Growth Pulses Evidenced by Small Animal Exploitation. *Science* 283:190–194.
- Stock, C. 1929. A Census of the Pleistocene Mammals of Rancho La Brea, Based on the Collections of the Los Angeles Museum. *Journal of Mammalogy* 10:281–289.
- Styles, B. W. 1981. *Faunal Exploitation and Resource Selection: Early Late Woodland Subsistence in the Lower Illinois Valley*. Northwestern University Archeological Program, Evanston, Illinois.
- Tedford, R. H. 1970. Principles and Practices of Mammalian Geochronology in North America. In *Proceedings of the North American Paleontological Convention*, edited by E. L. Yochelson, pp. 666–703. Allen Press, Lawrence, Kansas.
- Thomas, D. H. 1969. Great Basin Hunting Patterns: A Quantitative Method for Treating Faunal Remains. *American Antiquity* 34:392–401.
- Thomas, D. H. 1971. On Distinguishing Natural from Cultural Bone in Archaeological Sites. *American Antiquity* 36:366–371.
- Thomas, D. H., and D. Mayer. 1983. Behavioral Faunal Analysis of Selected Horizons. In *The Archaeology of Monitor Valley 2, Gatecliff Shelter*, edited by D. H. Thomas, pp. 353–391. Anthropological Papers of the American Museum of Natural History Vol. 59, No. 1. New York.
- Thompson, J. C. 2005. The Impact of Post-Depositional Processes on Bone Surface Modification Frequencies: A Corrective Strategy and Its Application to the Loiyangalani Site, Serengeti Plain, Tanzania. *Journal of Taphonomy* 3:67–89.
- Thornton, M., and J. Fee. 2001. Rodent Gnawing as a Taphonomic Agent: Implications for Archaeology. In *People and Wildlife in Northern North America: Essays in Honor of R. Dale Guthrie*, edited by S. C. Gerlach and M. S. Murray, pp. 300–306. British Archaeological Reports International Series 944. Oxford.
- Tipper, J. C. 1979. Rarefaction and Rarefaction – The Use and Abuse of a Method in Paleontology. *Paleobiology* 5:423–434.
- Todd, L. C. 1987. Taphonomy of the Horner II Bone Bed. In *The Horner Site: The Type Site of the Cody Cultural Complex*, edited by G. C. Frison and L. C. Todd, pp. 107–198. Academic Press, Orlando.
- Todd, L. C., and G. C. Frison. 1992. Reassembly of Bison Skeletons from the Horner Site: A Study in Anatomical Refitting. In *Piecing Together the Past: Applications of Refitting Studies in Archaeology*, edited by J. L. Hofman and J. G. Enloe, pp. 63–82. British Archaeological Reports International Series 578. Oxford.
- Todd, L. C., M. G. Hill, D. J. Rapson, and G. C. Frison. 1997. Cutmarks, Impacts, and Carnivores at the Casper Site Bison Bonebed. In *Proceedings of the 1993 Bone Modification Conference*,

- Hot Springs, South Dakota*, edited by L. D. Hannus, L. Rossum, and R. P. Winham, pp. 136–157. Occasional Paper No.1, Archaeology Laboratory, Augustana College, Sioux Falls, South Dakota.
- Todd, L. C., and D. J. Stanford. 1992. Application of Conjoined Bone Data to Site Structural Studies. In *Piecing Together the Past: Applications of Refitting Studies in Archaeology*, edited by J. L. Hofman and J. G. Enloe, pp. 21–35. British Archaeological Reports International Series 578. Oxford.
- Trapani, J., W. J. Sanders, J. C. Mitani, and A. Heard. 2006. Precision and Consistency of the Taphonomic Signature of Predation by Crowned Hawk-Eagles (*Stephanoaetus coronatus*) in Kibale National Park, Uganda. *Palaios* 21:114–131.
- Tuma, M. W. 2004. Middle to Late Archaic Period Changes in Terrestrial Resource Exploitation. *Journal of California and Great Basin Anthropology* 24:53–68.
- Turner, A. 1980. Minimum Number Estimation Offers Minimal Insight in Faunal Analysis. *Ossa* 7:199–201.
- Turner, A. 1983. The Quantification of Relative Abundances in Fossil and Subfossil Bone Assemblages. *Annals of the Transvaal Museum* 33:311–321.
- Turner, A. 1984. Behavioural Inferences Based on Bone Assemblages from Archaeological Sites. In *Frontiers: Southern African Archaeology Today*, edited by M. J. Hall, G. Aver, D. M. Avery, M. L. Wilson, and A. J. B. Humphreys, pp. 363–366. British Archaeological Reports International Series No. 207. Oxford.
- Turner, A. 1989. Sample Selection, Schlepp Effects and Scavenging: The Implications of Partial Recovery for Interpretations of the Terrestrial Mammal Assemblage from Klasies River Mouth. *Journal of Archaeological Science* 16:1–12.
- Turner, A., and N. R. J. Fieller. 1985. Considerations of Minimum Numbers: A Response to Horton. *Journal of Archaeological Science* 12:477–483.
- Uerpmann, H. P. 1973. Animal Bone Finds and Economic Archaeology: A Critical Study of Osteoarchaeological Method. *World Archaeology* 4:307–322.
- Vale, D., and R. H. Gargett. 2002. Size Matters: 3-mm Sieves Do Not Increase Richness in a Fishbone Assemblage from Arrawarra I, An Aboriginal Australian Shell Midden on the Mid-north Coast of New South Wales, Australia. *Journal of Archaeological Science* 29:57–63.
- Valensi, P. 2000. The Archaeozoology of Lazaret Cave (Nice, France). *International Journal of Osteoarchaeology* 10:357–367.
- van Es, L. 1995. Faunal Remains from Tell Abu Sarbut, A Preliminary Report. In *Archaeozoology of the Near East II*, edited by H. Buitenhuis and H.-P. Uerpmann, pp. 88–96. Backhuys, Leiden, The Netherlands.
- Van Valen, L. 1964. Relative Abundance of Species in Some Fossil Mammal Faunas. *American Naturalist* 98:109–116.
- Vasileiadou, K., J. J. Hooker, and M. E. Collinson. 2007. Quantification and Age Structure of Semi-Hypsodont Extinct Rodent Populations. *Journal of Taphonomy* 5:15–41.
- Vermeij, G. J., and G. S. Herbert. 2004. Measuring Relative Abundance in Fossil and Living Assemblages. *Paleobiology* 30:1–4.

- Vitousek, P. M., H. A. Mooney, J. Lubchenco, and J. M. Melillo. 1997. Human Domination of Earth's Ecosystems. *Science* 277:494–499.
- Voorhies, M. R. 1969. *Taphonomy and Population Dynamics of an Early Pliocene Vertebrate Fauna, Knox County, Nebraska*. University of Wyoming Contributions to Geology Special Paper No. 1. Laramie.
- Voorhies, M. R. 1970. Sampling Difficulties in Reconstructing Late Tertiary Mammalian Communities. In *Proceedings of the North American Paleontological Convention*, edited by E. L. Yochelson, pp. 454–468. Allen Press, Lawrence, Kansas.
- Waguespack, N. M. 2002. Caribou Sharing and Storage: Refitting the Palangana Site. *Journal of Anthropological Archaeology* 21:396–417.
- Ward, D. J. 1984. Collecting Isolated Microvertebrate Fossils. *Zoological Journal of the Linnean Society* 82:245–259.
- Watson, J. P. N. 1972. Fragmentation Analysis of Animal Bone Samples from Archaeological Sites. *Archaeometry* 14:221–228.
- Weinstock, J. 1995. Some Bone Remains from Carthago, 1991 Excavation Season. In *Archaeozoology of the Near East II*, edited by H. Buitenhuis and H.-P. Uerpmann, pp. 113–118. Backhuys, Leiden, The Netherlands.
- Weissbrod, L., T. Dayan, D. Kaufman, and M. Weinstein-Evron. 2005. Micromammal Taphonomy of el-Wad Terrace, Mount Carmel, Israel: Distinguishing Cultural from Natural Depositional Agents in the Late Natufian. *Journal of Archaeological Science* 32:1–17.
- White, T. E. 1952. Observations on the Butchering Technique of Some Aboriginal Peoples: I. *American Antiquity* 17:337–338.
- White, T. E. 1953a. A Method of Calculating the Dietary Percentage of Various Food Animals Utilized by Aboriginal Peoples. *American Antiquity* 19:396–398.
- White, T. E. 1953b. Studying Osteological Material. *Plains Archaeological Conference News Letter* 6:58–66.
- White, T. E. 1955. Observations on the Butchering Techniques of Some Aboriginal Peoples Numbers 7, 8, and 9. *American Antiquity* 21:170–178.
- White, T. E. 1956. The Study of Osteological Materials in the Plains. *American Antiquity* 21:401–404.
- Whittaker, R. H. 1972. Evolution and Measurement of Species Diversity. *Taxon* 21:213–251.
- Whittaker, R. H. 1977. Evolution of Species Diversity in Land Communities. *Evolutionary Biology* 10:1–67.
- Wild, C. J., and R. K. Nichol. 1983a. A Note on Rolf W. Lie's Approach to Estimating Minimum Numbers from Osteological Samples. *Norwegian Archaeological Review* 16:45–48.
- Wild, C. J., and R. K. Nichol. 1983b. Estimation of the Original Number of Individuals from Paired Bone Counts Using Estimators of the Krantz Type. *Journal of Field Archaeology* 10:337–344.
- Williams, C. B. 1949. Jaccard's Generic Coefficient and Coefficient of Floral Community, in Relation to the Logarithmic Series and the Index of Diversity. *Annals of Botany* 49:53–58.
- Winder, N. P. 1991. How Many Bones Make Five? The Art and Science of Guesstimation in Archaeozoology. *International Journal of Osteoarchaeology* 1:111–126.

- Wing, E. S., and A. B. Brown. 1979. *Paleonutrition: Method and Theory in Prehistoric Foodways*. Academic Press, New York.
- Witt, A. 1960. Length and Weight of Ancient Freshwater Drum, *Aplodinotus grunniens*, Calculated from Otoliths Found in Indian Middens. *Copeia* 1960:181–185.
- Wolda, H. 1981. Similarity Indices, Sample Size and Diversity. *Oecologia* 50:296–302.
- Wolff, R. G. 1973. Hydrodynamic Sorting and Ecology of a Pleistocene Mammalian Assemblage from California (U.S.A.). *Palaeogeography Palaeoclimatology Palaeoecology* 13:91–101.
- Wolff, R. G. 1975. Sampling and Sample Size in Ecological Analysis of Fossil Mammals. *Paleobiology* 1:195–204.
- Wolverton, S. 2002. NISP:MNE and %Whole in Analysis of Prehistoric Carcass Exploitation. *North American Archaeologist* 23:85–100.
- Wolverton, S. 2005. The Effects of the Hypsithermal on Prehistoric Foraging Efficiency in Missouri. *American Antiquity* 70:91–106.
- Wood, W. R. 1962. Notes on the Bison Bone from the Paul Brave, Huff, and Demery Sites (Oahe Reservoir). *Plains Anthropologist* 7:201–204.
- Wood, W. R. 1968. Mississippian Hunting and Butchering Patterns: Bone from the Vista Shelter, 23SR-20, Missouri. *American Antiquity* 33:170–179.
- Wright, D. H., B. D. Patterson, G. M. Mikkelsen, A. Cutler, and W. Atmar. 1998. A Comparative Analysis of Nested Subset Patterns of Species Composition. *Oecologia* 113:1–20.
- Wroe, S., T. Myers, F. Seebacher, B. Kear, A. Gillespie, M. Crowther, and S. Salisbury. 2003. An Alternative Method for Predicting Body Mass: The Case of the Pleistocene Marsupial Lion. *Paleobiology* 29:403–411.
- Zar, J. H. 1996. *Biostatistical Analysis*, 3rd ed. Prentice Hall, Upper Saddle River, New Jersey.
- Ziegler, A. C. 1965. The Role of Faunal Remains in Archeological Investigations. In *Symposium on Central California Archeology*, edited by F. Curtis, pp. 47–75. Sacramento Anthropological Society Papers No. 3. Sacramento, California.
- Ziegler, A. C. 1973. *Inference from Prehistoric Faunal Remains*. Addison-Wesley Module in Anthropology No. 43. Reading, Massachusetts.
- Zohar, I., and M. Belmaker. 2003. Size Does Matter: Methodological Comments on Sieve Size and Species Richness in Fishbone Assemblages. *Journal of Archaeological Science* 32:635–641.



# INDEX

---

- Abe, Y., 286–289
- Abundance
- absolute, 13
  - relative, 3, 14, 32
- Accuracy, 13, 108
- Actual number of individuals (ANI), 39, 71
- Adams, B. J., 127, 130
- Adams, W. R., 39–41, 57–58, 119
- Adams's dilemma, 58, 70, 78, 80
- Aggregate, 16, 70, 77, 78, 113, 123, 133
  - defined, 67
- Aggregation, 57–66, 223, 227, 262
- Allometry, 94, 108
- Andrews, P., 238
- Archaeofauna, 17
- Assemblage
  - defined, 16
  - fossil, 68
  - identified, 23, 141
- Atmar, W., 168–169
- Badgley, C., 4
- Barrett, J. H., 101, 102
- Bayham, F. E., 205
- Behrensmeyer, A. K., 267, 270–271
- Bergmann's rule, 109
- Betts, M. W., 92–93
- Binford, L. R., 92, 99, 104, 217, 218, 230, 233–238, 280
- Biocoenose, 22–25, 127, 173
- Biodiversity, 179
- Biomass, 84–90, 93–95, 96, 98, 99–113, 138, 139, 172
  - as a ratio scale measure, 86
  - as an ordinal scale measure, 107, 108
- Bobrowsky, P. T., 51, 52, 302–305
- Brain, C. K., 237, 275
- Braun, D. R., 287, 292
- Breitburg, E., 54
- Broad spectrum, 179
- Broughton, J. M., 137, 205, 208, 258, 303
- Brown, E. R., 87
- Bunn, H. T., 218, 219–220, 280, 285
- Buzas, M. A., 163
- Cain, C. R., 275
- Cannon, M. D., 158, 200–201
- Casteel, R. W., 3, 50–52, 94, 96–99, 300, 302, 305
- Cathlapotle, 18–20, 52, 53, 61, 73, 85, 87, 89, 90, 135, 146–148, 150, 175, 180, 186–187, 189, 192, 195
- Chaplin, R. E., 40, 94–95, 96, 99, 121–124, 129, 217
- Clason, A. T., 83
- Cleland, C. E., 179
- Closed array, 14
- Cochran's test for linear trends, 201, 204, 205, 211
- Community, 163, 173, 178, 201, 210
  - defined, 22
  - distal, 242, 244, 248
  - proximal, 241, 244, 248

- Cook, S. F., 94
- Corrected number of specimens per individual (CSI), 242, 244
- Cruz-Uribe, K., 11, 51–52, 218, 250, 302, 303
- Dean, R. M., 118
- Diversity, 174  
defined, 173, 175, 209  
kinds of, 173
- Domínguez-Rodrigo, M., 296
- Ducos, P., 50
- Dunnell, R. C., 179
- Efremov, I. A., 7, 264
- Egeland, C. P., 289
- Emerson, T. E., 111, 138
- Emeryville Shellmound, 205
- Enloe, J. G., 256
- Estimate  
defined, 8
- Evenness, 177, 192, 194–198, 201, 202  
as an ordinal scale measure, 201–202  
defined, 175
- Fagerstrom, J. A., 68
- Faunule, 22, 68
- Fidelity studies, 24, 211, 298
- Fieller, N. R. J., 126–128
- Flannery, K. V., 179
- FLK Zinjanthropus, 220, 269
- Ford, J. A., 68
- Fragmentation  
extent of, 251  
intensity of, 43, 252
- Frey, C. J., 224, 255
- Gautier, A., 37
- Gifford-Gonzalez, D. P., 281, 282
- Gilbert, B. M., 282
- Grayson, D. K., 3, 4, 11, 25, 29, 49, 50, 51, 52, 54, 57–58, 64, 65, 67, 71–77, 80, 81, 128, 139, 140, 155, 162, 199, 208, 224, 243, 255, 299, 300, 302, 305
- Guilday, J. E., 113, 282
- Gust, S. M., 92
- Guthrie, R. D., 85–87, 88, 89, 93, 108
- Hayek, L.-A., 163
- Herbert, G. G., 31–32
- Hesse, B., 11, 51, 52, 305
- Heterogeneity, 176, 177, 201, 202  
as an ordinal scale measure, 201–202  
defined, 176
- Holland, S., 166
- Holtzman, R. C., 134
- Homestead Cave, 181, 207
- Howard, H., 1, 11, 21, 38, 39, 68, 119
- Hudson, J., 54
- Identifiable specimens  
defined, 6
- Interdependence, 17, 22, 47, 273
- Interdependence of identified specimens, 30, 37–39, 44, 54, 57, 71, 73, 77, 109, 112, 173, 222  
interaggregate, 58, 69
- Jaccard index, 186, 187, 189
- Jackson, H. E., 99
- Jones, E. L., 202
- Kim, S. Y., 255
- Kintigh, K. W., 162, 163, 171, 185
- Klein, R. G., 11, 43, 44, 51–52, 218, 250, 302, 303
- Kobeh Cave, 255
- Konigsberg, L. W., 127, 130
- Kowalewski, M., 32, 296–297
- Krantz, G. S., 129
- Kroll, E. M., 219, 280
- Landon, D. B., 74
- Le Flageolet I, 183
- Leonard, R. D., 143, 171
- Lincoln/Petersen index, 84, 123, 124–129, 133, 139
- Local fauna, 68
- Lupo, K. D., 296

- Maltby, J. M., 282
- Marean, C. W., 11, 219, 221, 255, 277
- Marshall, F., 43, 253, 258, 300
- Maximum likelihood, analysis of bone counts by (ABCML), 136
- Measurement  
defined, 8  
derived, 12  
fundamental, 12, 13
- Meat weight, 41, 84, 89–94, 96, 99, 102, 106, 108, 113, 139  
as a ratio scale measure, 92  
as an ordinal scale measure, 108
- Meier site, 17–20, 52, 53, 87, 144–148, 150, 151, 165, 166, 175, 180, 186, 187, 189, 192, 195, 224–227, 248
- Milo, R. G., 285
- Minimum animal units (MAU), 234, 245
- Minimum number of elements (MNE), 214, 217, 218  
as a derived measurement, 218, 263  
as a ratio scale measure, 228  
as an ordinal scale measure, 222–227, 258–259, 263  
defined, 215, 218, 220
- Minimum number of individuals (MNI)  
as a derived measurement, 12, 46, 79, 214  
as a ratio scale measure, 48, 72, 78  
as an ordinal scale measure, 48, 72, 78  
defined, 38  
relationship to NISP, 49
- Mollhagen, T. R., 42
- Morrison, D. A., 74
- Nagaoka, L., 202
- Needs-Howarth, S., 106, 108
- Nestedness, 167–170, 191
- Ngamuriak, 259
- Nichol, R. K., 132
- Number of identified specimens (NISP)  
as a fundamental measurement, 28, 79, 263  
as a ratio scale measure, 72, 78  
as an ordinal scale measure, 36, 72, 79, 173, 190  
defined, 27  
relationship to MNI, 49
- Number of specimens (NSP), 266  
as a fundamental measurement, 266  
defined, 27
- Number of taxa (NTAXA). *See also Richness*, 143
- Number of unidentified specimens (NUSP), 266
- O'Connell, J. F., 296
- Olduvai Gorge, 268
- Olsen, S. L., 282
- Olszewski, T., 178
- Orchard, T. J., 137, 138
- Overlap, anatomical, 39, 214, 215, 219, 220, 221
- Patterson, B. D., 168–169
- Payne, S., 67, 152, 153, 154
- Perkins, D., Jr., 21, 37
- Pilgram, T., 43, 253, 258, 300
- Plug, C., 47
- Pobiner, B. L., 287, 292
- Potts, R., 220, 270
- Prolom II Cave, 256
- Purdue, J. R., 111
- Quitmyer, I. R., 105
- Rancho La Brea, 1, 2, 8, 13, 21, 24, 25, 29, 38, 175, 215
- Rarefaction, 160, 165, 166, 167, 170, 180, 189–191
- Reed, C. A., 94
- Reitz, E. J., 68, 103–105, 244, 299
- Reliability, 12
- Rhode, D., 162
- Richness, 159, 174, 177, 179–185, 201, 202  
as a ratio scale measure, 179  
as an ordinal scale measure, 201–202  
defined, 175
- Richness, taxonomic. *See Number of taxa (NTAXA)*
- Ringrose, T. J., 44, 67, 127, 261, 300, 301

- Rogers, A. R., 136–137  
 Rose, J., 281
- Sanders, H. L., 162  
 Scale of measurement  
     interval, 10  
     nominal, 9, 77  
     ordinal, 9, 77, 205  
     ratio, 10, 190, 205  
 Schulz, P. D., 92  
 Shannon index of evenness, 195, 246, 248  
 Shannon-Wiener index, 192, 195  
 Shipman, P., 281, 282  
 Shotwell, J. A., 4, 30–32, 68, 134–136, 210, 241–248  
 Sibudu Cave, 275  
 Simpson's index  
     reciprocal of, 196–197  
 Skeletal element  
     defined, 5  
 Skeletal part, 5  
 Skeletal portion, 5  
 Smith, B. D., 89, 90, 91  
 Sorenson index, 186, 187, 189  
 Sorenson's quantitative index, 189  
 Sparks, B. W., 152  
 Species-area relationship, 180  
 Specimen  
     defined, 5  
 Spencer, L. M., 219  
 Stahl, P. W., 91  
 Stewart, F. L., 91  
 Stock, C., 1–2, 8, 11, 13, 21, 24, 25, 27, 29, 38, 39, 41, 42, 68, 119, 175, 215, 217  
 Styles, B. W., 114  
 Surface area solution, 277  
 Survivorship, 238
- Taphocoenose, 23–24, 127  
 Taphonomic signature, 264  
 Taphonomy, 264  
     defined, 7  
 Thanatocoenose, 22–24, 127, 173  
 Thomas, D. H., 154–155, 158, 244, 248  
 Time averaging, 211, 212  
 Tipper, J. C., 160  
 Todd, L. C., 129  
 Treganza, A. E., 94  
 Turner, A., 22, 25, 126–128
- Ubiquity, 84, 114–119, 140  
 Uerpmann, H. P., 65, 102, 108
- Valensi, P., 69  
 Validity, 12, 13, 128, 140, 276, 277, 289  
 Van Valen, L., 31, 32  
 Variable  
     continuous, 9  
     discontinuous, 9  
     measured, 11, 24  
     target, 11, 24  
 Vermeij, G. J., 31–32  
 Voorhies, M. R., 4, 216–217, 230  
 Voorhies' Groups, 217
- Wapnish, P., 11  
 Watson, J. P. N., 152, 286  
 Weighted abundance of elements (WAE), 134–136  
 White, T. E., 39, 41, 43, 46, 85, 89–93, 94, 95, 102, 108, 119–120, 121, 140, 217, 229–232, 235, 261, 300  
 Wild, C. J., 132  
 Winder, N. P., 128  
 Wing, E. S., 68, 244, 299  
 Wolff, R. G., 144, 184