HISTORIA ANIMALIUM EX OSSIBUS

Beiträge zur Paläoanatomie,
Archäologie, Ägyptologie, Ethnologie
und
Geschichte der Tiermedizin

FESTSCHRIFT
FÜR
ANGELA VON DEN DRIESCH
zum 65. Geburtstag

herausgegeben von
Cornelia Becker, Henriette Manhart, Joris Peters
und Jörg Schibler

Verlag Marie Leidorf GmbH · Rahden/Westf.
1999
The use of size index scaling techniques for research on archaeozoological collections from the Middle East

Richard H. Meadow, Cambridge MA

Summary

Beginning with the work of Pierre Ducos published in 1968, techniques for scaling animal bone measurements from archaeological sites in the Middle East have been employed to investigate body size changes in sheep, goat, cattle, and pigs. Four more or less related algorithms have been developed to permit investigators to combine on a single graph and analyze measurements from different skeletal parts. These are: the SI (Size Index) first introduced by Ducos, the RSI (Relative Size Index) and VSI (Variability Size Index) both created by Hans-Peter Uerpmann, and the LSI (Logarithm Size Index) first employed by the author. In this essay, each of the four scaling techniques is defined, and their relationships to each other demonstrated. Assumptions underlying their use and associated problems are discussed, and recommendations are made with respect to their use in the future. It is concluded that size indices have a role to play, particularly with respect to small collections, but that they be employed as only one of a number of techniques useful for investigating animal exploitation patterns in the past.

Keywords: Size index, Dimensional scaling, Animal domestication, Measurements, Archaeozoology.

Introduction

In their 1997 article on the fauna from the Prepottery Neolithic site of ‘Ain Ghazal in Jordan, Angela von den Driesch and Ursula Wodtke employ a technique of scaling animal bone measurements that was first published by me in 1981. Now generally termed the “log size index” (LSI) technique, it has become widely used by faunal analysts working on collections of mammal bones from the Middle East. The technique has been employed to investigate variability in animal size through time and across space when analytical units of interest contain only small numbers of measurable skeletal parts (e.g., Grigson 1989; Helmer 1989; 1992; Uerpmann & Uerpmann 1994). Other scaling (size index) techniques employed in interpreting metrical data from sites in the Middle East were developed and first published by Pierre Ducos in 1968 and by Hans-Peter Uerpmann in 1979. All such approaches to investigating metrical data have their strengths and weaknesses, both intrinsic and in the way that they are applied by investigators. In this essay I review the most commonly used scaling techniques and comment upon their relationships and upon the underlying assumptions of and problems in their use. I also make some suggestions about how they might be better employed in the future. In so doing I extend observations that have been made in the literature by Ducos (1991), Meadow (1991, 1993), Uerpmann and Uerpmann (1994), and Ducos and Horwitz (1997).

The scaling techniques and their first use

Introduction

Faunal assemblages from archaeological sites generally contain relatively few measurable specimens when compared to the number of identified fragments (By “measurable” I mean skeletal parts on which a standard set of anatomically meaningful dimensions can be taken, e.g., following the definitions of von den Driesch 1976). When the area excavated at a site is restricted or when a large collection is subdivided into chronological or spatial subsets, the number of comparable measurements from a single skeletal part in a single analytical unit may be very small indeed. For example, there may be
fewer than a half dozen each of measurable distal humeri, proximal radii, distal radii, and so on for any given taxon. Under such circumstances the accuracy of the usual measures of central tendency (e.g., mean and median) and of dispersion (e.g., standard deviation) for each dimension are likely to be affected by "sampling error."

**Ducos and the SI**

More than thirty years ago, Pierre Ducos (1968, 137) recognized this problem and suggested:

"Pour augmenter les possibilités de comparaison, il est nécessaire d’attribuer à chaque ossement un indice qui, de façon acceptable, soit proportionnel à la taille de l’individu auquel il a appartenu, de sorte que les indices ainsi calculés soient comparables entre eux, quelle que soit la partie de squelette à laquelle ils se rapportent respectivement."

[In order to increase the possibilities for comparison, it is necessary to allocate to each bone an index that, in an acceptable fashion, is proportional to the size of the individual to which it belonged so that the indices so calculated are comparable no matter to what part of the skeleton they respectively belong.]

His solution was to calculate averages for each of several selected skeletal dimensions from one of his assemblages and use those averages as standards against which to compare the individual measurements from other assemblages. Measurements of other specimens could be larger than, the same as, or smaller than that of the standard, but all could be related to the standard in comparable fashion by dividing the one by the other according to the following formula (explicitly stated by Ducos only in 1991, p. 161):

\[
\text{indice de taille} = \frac{x}{m}, \text{pour le } p-ième \text{ élément de la catégorie } i.\]

Here I generalize the formula as $SI$ (size index) $= \frac{x}{m}$, with $x$ being the dimension of the archaeological specimen and $m$ being the standard whether this be a mean (as for Ducos) or a measurement from a single animal. When the archaeological specimen has the same dimension as the standard, the size index will be equal to 1, with specimens larger than the standard being greater than 1 and specimens smaller than the standard being less than 1 but greater than 0. For each genus, Ducos plotted the ratios so obtained together on the same graph by site and/or by archaeological unit. On his graphs, Ducos (1968, e.g., fig. 22) assigned the x-axis to the ratio (or index) and the y-axis to the number of specimens falling within a ratio interval (e.g., 1.001-1.025, 1.026-1.050, 1.051-1.075, 1.076-1.100, etc.), although the interval was not explicitly defined. He inspected the resulting graphs and compared them visually for differences in bone size distributions.

The size index technique of Ducos (1968) dropped from sight for many years. He did not use it again for more than two decades (Ducos 1991), and other investigators may not have picked up on it because Ducos did not present his approach very explicitly in the original publication (no formula was given). It is also possible that the approach was overlooked because Ducos employed the technique on the bones of cattle ($Bos$) and pig ($Sus$) but not on those of sheep ($Ovis$) or goat ($Capra$). During the 1970s and 1980s, far greater attention came to be paid to the latter two forms than to the former because of the dominance of sheep and goat bones in Near Eastern faunal collections.

**Uerpmann and the RSI and VSI**

For examining size change through time in sheep and goat, Hans-Peter Uerpmann independently came up with a related approach for combining measurements from different skeletal parts on the same graphical axis. This he first published as Appendix 2 of his 1979 volume "Probleme der Neolithisierung des Mittelmeerraums" under the heading "Der Größenvergleich von Tierknochenfunden mit Hilfe von Größenindices." Like Ducos, Uerpmann took the approach of relating dimensions of archaeological specimens to those of a standard. His standards, however, were not means derived from
The use of size index scaling techniques

In order to evaluate the differences between the dimensions of the "standard individual" (or the "standard animal") and the archaeological specimens, Uerpmann decided to use a scale of 100 units, with those 100 units being equal to six coefficients of variation, three on either side of the standard dimension. Given a likely coefficient of variation of between 9% and 15% for sheep and goat from the same population (higher in the latter than in the former), a figure of 12% was arbitrarily selected for use. Thus:

\[ v = \pm 36b/100 \] or \[ v = 72b/100 \], where \( v \) = the variation and \( b \) = the standard dimension. For each unit \( e \) of the 100: \( e = v/100 = 72b/10,000 \). On the scale of 100, 50 was taken to be the mid-point, i.e., equal to the standard dimension. Thus the size index comes out as:

\[ i = [(m-b)/e]+50 \]

for any measurement \( m \) of an archaeological specimen when compared to the same dimension of the standard individual \( b \).

Carrying out the substitution for \( e \), the final formula is:

\[ i = [(1250m/b)-800]/9. \]

Using the same abbreviations as Ducos employed (\( x = \) the archaeological specimen, \( m = \) the corresponding standard dimension), the formula becomes:

\[ RSI \text{ (relative size index)} = [(1250x/m)-800]/9, \] with the standard dimension being set, by this formula, at 50 (i.e., when \( x = m, RSI = 50 \)). It is also possible to subtract 50 from the RSI thus placing the standard dimension at 0.

Like Ducos (1968), Uerpmann (1979, Diagramms 2-5) plotted his RSI data on graphs, assigning the x-axis to the index. Instead of using counts of specimens per interval, however, Uerpmann calculated the percentage of the total number of specimens that fell within each 5-unit interval. As in the case of Ducos, the exact range used by Uerpmann is not defined; for example, if an RSI is 50 exactly, it is not clear whether it was grouped with specimens in the 45-50 interval or in the 50-55 interval. In addition to this use of frequency histograms, Uerpmann placed arrows on his graphs to indicate the location of the median RSI for each analytical unit (in his case, time periods). In the accompanying text he emphasized that it is the change in the location of the median together with the change in the location of the smallest RSI category that needs to be examined when evaluating possible size diminution in sheep and goats through time.

In his 1982 article on the “Faunal remains from Shams ed-Din Tannira, a Halafian site in northern Syria,” Uerpmann developed another size index to deal with equids and Gazella for which means and standard deviations were available for the standard animals. Instead of using the ratio of archaeological specimen to standard (as in SI, RSI, or LSI), he developed a function expressing difference from the standard in terms of numbers of standard deviations. Here

\[ VSI \text{ (variability size index)} = 50(x-m)/2s, \] where \( s \) is equal to the standard deviation of the mean \( (m) \) of the measurement series (the “standard”) to which an archaeological measurement is being compared.

Instead of assuming or estimating a coefficient of variation as in the RSI, the actual standard deviation of the mean for each chosen dimension of each selected skeletal part in a population (modern or ancient) is used. For example, if the equid remains from Mureybit (Ducos 1978) are chosen to provide an equid standard, the mean LG of the Scapula is 44.1 mm with a standard deviation of 1.98,
the mean BT of the Humerus is 60.0 mm with a standard deviation of 1.77, etc. In this case the standard dimension \( m \) will fall at 0 on the VSI diagram and not at 50 as on the RSI diagram. As for the variable \( 2s \), the constant is not fixed and can be \( 1s \) or \( 2s \) or \( 3s \) or \( 4s \) and so on depending upon where the investigator wishes to set -50 and +50 in relation to the standard.

In the 1982 paper, Uerpmann plotted his VSI values using cumulative frequency graphs instead of histograms (Uerpmann 1982, fig. 1 for equids and fig. 3 for Gazella, and also fig. 7 for goats using RSI). Cumulative frequency graphs, in which individual size indices are plotted in increasing or decreasing order, permit the investigator to identify visually the degree of normalcy of, and the presence of any gaps in, the distribution. A normal distribution is somewhat S-shaped while discontinuities are revealed by unusually large gaps in the size index values.

**Meadow and the LSI**

While compiling measurement data from Mehrgarh in 1979 (eventually published in 1981), I was fully aware of Uerpmann’s RSI technique. However, I found the RSI to be cumbersome to calculate. Being ignorant of the approach of Ducos, and at that point not recognizing that all that the constants actually do in the RSI formula is to scale the results of the simple ratio \( x/m \), I turned to the “ratio diagram” technique of George Gaylord Simpson first published in 1941 (and subsequently summarized in Simpson, Roe & Lewontin 1960, 356ff.). This technique was developed in order to “compare graphically the relative rather than the absolute dimensions of a number of animals or groups of animals” (ibid., 356).

“The first step in constructing the diagram is to convert all the measurements into logarithms (it does not matter to what base). One specimen or group of specimens is then chosen as the standard of comparison. For each dimension the difference between the logarithmic value of the standard and each of the other specimens or groups is calculated (ibid., 357).” In other words,

\[ d = \log X - \log \text{standard} \]

Again, using the same abbreviations that Ducos employed, \( LSI \) (log size index) = \( \log x - \log m \), with specimens larger than the standard having positive values and those smaller than the standard having negative values. An important feature of using logarithms was that “the horizontal distance between any two points in the [ratio] diagram is proportional to the ratio of the dimensions of the two animals. This is because \( \log X - \log Y = \log(X/Y) \)” (ibid., 357). Thus, in ratio diagrams where the investigator is trying to visually compare different dimensions in a single element, it makes a good deal of sense to use logarithms instead of a non-logarithmic ratio technique like that of Ducos \( (x/m) \) (Another reason for using logarithms was that, before desk-top computers and even electronic calculators, it was easier to look up the logarithms in a table and to subtract one from another than it was to do long division by hand (Simpson 1941, 24).

My first use of the LSI (or log difference) technique was to investigate the possibility of size diminution in cattle, sheep, and goat through the course of the aceramic Neolithic at the site of Mehrgarh, which is located in the east central part of Pakistani Baluchistan. In the graphs (Meadow 1981, figs. 5, 6 and also in Meadow 1983, fig. 3) I did not employ histograms but instead individually plotted each log difference using an abbreviation for the skeletal part represented. In this way it was possible to evaluate the range of variability for each element, seeing which ones were more variable and which ones less so. Following Uerpmann, I also plotted the median point for each chronological unit. As standards, I used those defined by Uerpmann (1979) for sheep and goat and another for \( Bos \text{ indicus} \) that I measured. These were published in the article (Meadow 1981, table 9) and included length, breadth, and depth dimensions. Subsequently, I have employed histograms (e.g., Meadow 1984) and other techniques of illustrating size variability, including one similar to the cumulative frequency technique of Uerpmann (1982) in which individual log size indices are plotted in increasing or decreasing order (e.g., Meadow 1991; 1993).
The use of size index scaling techniques

Summary of the scaling techniques

In sum, the different scaling techniques discussed above are:

\[ SI \text{ (size index)} = \frac{x}{m} \]
\[ LSI \text{ (log size index)} = \log(x - \log m) = \log(x/m) \]
\[ RSI \text{ (relative size index)} = \frac{[(1250x/m)-800]}{9} \]
\[ VSI \text{ (variability size index)} = \frac{50(x-m)}{2s} \]

where \( x \) equals the dimension of the archaeological specimen and \( m \) equals the corresponding dimension of the “standard animal” or “standard population.” For calculating VSI, a series of dimensions that are averaged is required. Here \( s \) equals the standard deviation of the measurement series for the individual skeletal part to which the archaeological specimen is being compared. In the case of both the RSI and the VSI, one can vary the constants depending upon how one wishes to scale the results.

Relationships between the different size index techniques

Figures 1 to 6 show the relationships between the different size indices. Two hypothetical series of measurements are plotted. These are:

1) a data series from 13 to 32 mm in 0.5 mm increments. Indices are calculated using an arbitrary standard dimension of 22.8, with a coefficient of variation set at 12 %, which provides a standard deviation of 2.736. The symbols used on the graphs are small squares;

2) a data series from 17 to 36 mm in 0.5 mm increments. Indices are calculated using an arbitrary standard dimension of 26.6, with a coefficient of variation set at 9 %, which provides a standard deviation of 2.394. The symbols used on the graphs are X’s.

Figure 1 shows the relationship between LSI and SI. LSI is calculated using base 10 logarithms in all cases, although base e logarithms could have been used as well. The two curves are congruent, and the relationship between the two indices is curvilinear. Distances on the LSI axis between two points increase as the SI values get smaller. This pattern simply reflects the fact that \( LSI = \log SI \).

The same phenomenon can be seen in figure 2 and in figure 3. Figure 2 differs from figure 1 only in the scale used for the y-axis. In figure 3, however, the two curves are not congruent, but have different slopes. This reflects the larger coefficient of variation of the series represented by the squares (12 % instead of 9 %). The same phenomenon can be seen in figures 5 and 6.

Figures 4, 5, and 6 demonstrate that the relationships among SI, RSI, and VSI are linear with no increase or decrease between points as the index values get smaller or larger. Figure 4 underlines the fact that RSI is only a linear modification of SI that serves to change the units employed on the...
Fig. 3 The Logarithm Size Indices (LSI) for 78 values plotted against the Variability Size Indices (VSI) for those same values (meaning of symbols see text).

Fig. 4 The Size Indices (SI) for 78 values plotted against the Relative Size Indices (RSI) for those same values (meaning of symbols see text).

Fig. 5 The Size Indices (SI) for 78 values plotted against the Variability Size Indices (VSI) for those same values (meaning of symbols see text).

Fig. 6 The Relative Size Indices (RSI) for 78 values plotted against the Variability Size Indices (VSI) for those same values (meaning of symbols see text).

diagram \( RSI = \frac{[(1250SI) - 800]}{9} \). It is no surprise, therefore, that figures 5 and 6 are identical except for the scale used on the x-axis. The relationship between SI (and RSI) and VSI, while linear, is more complicated given that VSI is calculated by employing the actual variance of the sample providing each standard measurement.

Because VSI directly reflects differing skeletal part variability, the rank-order of a particular specimen in relation to other specimens can be different than if RSI (or SI or LSI) had been employed. This is demonstrated by figures 7 and 8, in which a set of 35 Bos bone breadth measurements from Bouqras (Buitenhuis 1988) are scaled using a standard calculated from the dimensions of bones of wild cattle from Denmark (Degerbøl 1970; here table 1). Figure 7 shows the ranked RSI distribution, with the individual skeletal parts labeled and numbered above the graph. If SI or LSI had been employed, the specimens would be arrayed in the same order. Figure 8 shows the ranked VSI distribution. The individual skeletal part labels and numbers refer to the same specimens as in figure 7. Of note is the fact that the rank order has changed. There are shifts toward the 0-line of VSI values calculated for elements with higher than average coefficients of variation such as the central-and-fourth tarsal. And there are corresponding shifts away from the 0-line of elements with lower than average coefficients of variation such as the proximal radius (table 1).

Is the VSI a “better” size index technique to use because it is more sensitive to differing variance among elements? A Spearman’s coefficient of rank-order correlation suggests that the rank-orders in which the bones are plotted using RSI (fig. 7) and VSI (fig. 8) are not significantly different
The use of size index scaling techniques

The basic underlying assumptions of, and problems with, size-index techniques

All the size index techniques discussed here depend upon comparing the dimensions of archaeological (or sometimes modern) specimens with the corresponding dimensions of a standard animal. This standard animal can be a single individual (in which case only the ratio techniques RSI, SI, and LSI can be used) or a “population” for which mean dimensions are calculated (in which case VSI can also be used if standard deviations for each mean are determined). In the case of the single animal, the same dimensions of the right and left sides are usually averaged. In all cases, each dimension or each average dimension of each skeletal part has a specific relationship to each other dimension in the skeleton of that single animal or “average animal”. It is these relationships within the standard that the investigator is using to compare the archaeological specimens to each other by means of relating each one to the standard. Thus when using a size index technique one assumes that the body proportions of the animals that provided their bones to the archaeological collection were the same as, or at least similar to, the body proportions of the standard animal, however defined.

There are all sorts of reasons why this might be a poor assumption. These relate to the inherent variability in the proportions of animals within a single population due to age (allometric growth), sex (dimorphism), or idiosyncratic factors related to the individual. They also relate to variations between local populations within the same species due to such environmental effects as those encapsu-
lated by Allen's and Bergmann's Rules as well as those due to specific stresses on specific populations. And, of course, such differences in proportion are likely to be even greater between species and, in the case of domestic animals, between breeds.

In an attempt to investigate what can happen when using different standards, I calculated VSIs and RSIs for a set of 142 breadth and depth measurements of gazelle bones that I studied from Neolithic (c. 7000-5500 calBC) deposits at Mehrgarh (Baluchistan, Pakistan; in the interests of space, since SI and LSI are directly related to RSI, I do not discuss them separately). These size indices were calculated using two standards: a *Gazella bennetti* standard (mean dimensions of two males and one female from eastern Iran) and a *Gazella gazella* standard (mean dimensions of two males and one female from Israel) (see table 2). Normally, one would never use for a standard the bones of a species different from that of the archaeological specimens. Sometimes, however, identifications cannot be made at the species level. Furthermore, modern wild forms may differ somewhat in proportions and in degree of variability from their presumed ancestors in the same region, and there may be considerable clinal variability even within contemporary species.

Figures 9 and 10 show the relationships between the RSIs (fig. 9) and VSIs (fig. 10) for *Gazella gazella* and *G. bennetti* as scatter plots. In neither case is the sample estimate of the correlation coefficient overwhelmingly high (r = 0.83 and 0.78, respectively). Figures 11 and 12 show the same data, this time grouped and displayed in sets of histograms. Here one can see reflected not only the differences in absolute size of the animals used as standards - *G. bennetti* being considerably smaller than *G. gazella* - but also the effects of differences in relationships between the dimensions of the various skeletal parts of those standards. Thus if a curve were fitted to the histograms based on *G. bennetti*, it would be more peaked (leptokurtic) than that based on *G. gazella*. This is particularly clear when the VSI technique is employed (fig. 12).

The principal purpose for which size index techniques have been employed is to document change in animal size over time. What is meant by size? This in and of itself is a problem. The original use of the various scaling techniques grouped indices calculated from length and breadth/depth measurements in the same graphs (e.g., Uerpmann 1979; Meadow 1981; 1984), and this practice continues
The use of size index scaling techniques

among many investigators who employ the techniques. Upon reflection, however, it has become clear
to both me (Meadow 1991) and Uerpmann (Uerpmann & Uerpmann 1994) that this is not a good idea.
Size is three dimensional, with bone length reflecting animal height and bone breadth and depth re­
flecting animal weight. To quote Uerpmann and Uerpmann (1994, 430f.):

“As a general rule, only those size indices should be grouped together that were calculated
from measurements depending on similar allometric parameters. There are three main groups of meas­
urements in mammal osteometry which have to be kept apart when calculating size indices. Skull
measurements should not be transformed into size indices at all, because they depend on brain size on
the one hand and tooth size on the other, which are influenced by independent evolutionary trends and
selective pressures. Length measurements of the long bones of the extremities depend on the height of
the animal and the length of its extremities. They are to some extent independent of the width and
depth measurements of the same bones and the measurements of the small bones of the extremities.
This last group of measurements depends largely on the weight of the animal. As the weight is among
the most important economic features of domesticates, and the starting point of these considerations,
we have restricted our calculations to this group of measurements.”

While bone breadths and depths reflect the economically important feature of body mass, bone
lengths - especially when taken together with breadths (slenderness index) - may be a particularly
useful reflection of both sex and breed differences in domestic animals. Given the nature of archaeo­
logical collections, however, it is usually only the phalanges and carpal and tarsal bones that survive
intact in large numbers. Of these, it is not yet entirely clear to what degree the length of any carpal or
tarsal might provide a good reflection of animal height. As for the phalanges, those of the fore and
hind limbs of ungulates can be of significantly different lengths, yet not always easily distinguishable.
Medial and lateral phalanges in artiodactyls can also be of significantly different breadths. Until these
problems are resolved, it seems preferable to keep all length dimensions strictly separate from those of
breadth and depth when employing size index techniques. It also might be best not use phalanx
breadths in artiodactyls unless they can be identified as fore or hind and as medial or lateral.

Even if one deals with length and breath/depth dimensions separately, however, differing fre­
quencies of different skeletal parts in two collections can potentially affect the distribution of size
indices. In figure 13, VSI values are plotted against RSI values for the Mehrgarh material using the
Gazella bennetti standard. The points group into two distinct linear clusters, reminiscent of the lines
seen in figure 6. Indeed, in figure 13 the more horizontal of the lines includes only those dimensions
with relatively high coefficients of variation in the standard (greater than v = 6.8: Ulna DPA, Pelvis
LA, Femur DC, Mc Bd, Mt Bd; see table 2). Those in the other line have lower coefficients of varia-
tion \((v = 4.25 \text{ or less}: \text{Scapula BG, Humerus BT, Radius Bp and Bd, Mc Bp, Tibia Bd, Talus Bd, Mt Bp})\). The second group (those with lower \(v\) values) contributes most of the high VSI values (36 out of 38 VSI values greater than 50). It also contributes most of the highest RSI values (greater than RSI = 70), although these are not so many (7 out of 10). A different distribution of skeletal part frequencies in the collection would be likely to lead to a different distribution of size indices, especially VSI in this instance, given the parameters of this particular \textit{Gazella bennetti} standard.

![Fig. 13 Scatter plot of breadth or depth dimensions of 142 \textit{Gazella} sp. bones from Mehrgarh (Baluchistan, Pakistan) having been converted to RSI and to VSI, using a \textit{Gazella bennetti} standard (see text and table 2).](image)

Another problem with using scaling techniques that combine indices for different skeletal parts on the same graph is the same as the problem that affects use of NISP (number of identified specimens) as a measure of taxonomic abundance, namely interdependence (e.g., Grayson 1984). One must assume that it does not really matter to the results that there may be multiple specimens from the same animal in the assemblage. This may not be a serious issue when scattered bones from across a highly differentiated site are grouped together into comparative units of convenience. It does, however, become an issue when body parts from the same contexts can be seen to articulate thus leading the analyst to suspect that whole skeletal segments may have been discarded together or intact. Under such circumstances does one include, for example, measurements for each of an articulating central+fourth tarsal, astragalus, and proximal metatarsal, or does one choose only one of them? Or does one average the three indices? A related problem is whether one should average indices calculated for different dimensions on the same skeletal part (e.g., Bd and Dd of the Humerus).

There is no clear answer to these questions. On the face of it, it seems obvious that one should average. But doing so may affect consistency and comparability, two of the most important goals of any analysis. An analyst may not find articulations when, in fact, they are there. The bones from those overlooked joints will thereby contribute relatively more to the series than will the articulating specimens. In a like manner, averaging two or more indices calculated for a single specimen provides a value that is not necessarily comparable to the value provided by either of the single dimensions because of individual variability. And for some specimens, it may only be possible to measure one dimension. This harks back to the problems of the assumption of uniformity in the relation of body proportions discussed above. As with NISP, in the end interdependence may not matter very much in most cases if the analyst is only trying to document trends. But in so saying, one must not over-interpret the results of size index manipulations.

A further problem with averaging is specific to its use with the LSI. As Ducos (1991) has pointed out, \(\frac{\text{Log}(x_1/m_1)+\text{Log}(x_2/m_2)}{2}\) does not equal \(\text{Log}(\frac{x_1/m_1+x_2/m_2}{2})\), i.e., the mean of logarithms of two values is not equal to the logarithm of the means of the same two values. Thus if the analyst averages two LSIs for a bone, the result will be different from that obtained by taking the logarithm of the average of two SIs. The value of the former (averaged LSIs) will be smaller than that of the latter (log average SIs) when the two SIs are not equal. Most, and perhaps all, researchers who have used LSIs and who have employed index averaging have taken the mean of the logarithms. Because LSIs plot out in a non-linear fashion, this approach, although never explicitly justified, is perhaps the better one.

Both Ducos (1991) and Uerpmann and Uerpmann (1994) have noted that using LSIs usually does not make a whole lot of difference in the results of a size index analysis, and furthermore,
"as animal growth is also not linear, but best described with exponential allometric parameters, this non-linearity of the LSI is only impractical, not a real fault. When applied to the remains of adult mammals, the difference between minima and maxima of normal ranges of variation is so small that the non-linearity can usually be neglected, as long as LSIs are calculated from individual find-measurements. It cannot be neglected, however, if the find-measurements have already been transformed into statistical parameters such as means or standard deviations. The algebraic mean of an LSI-distribution is not the logarithm of the mean value of the measurement-distribution! Therefore, LSIs may be calculated only from individual measurements, and their statistical parameters may not be translated back to be compared with mean values or standard deviations, etc., of linear measurements. LSIs can be used only for comparison with other LSIs, but if this rule is obeyed, they provide useful results" (Uerpmann & Uerpmann 1994, 429f.).

In addition, as Ducos (1991) has pointed out, the use of statistical tests relating to normal (Gaussian) distributions is technically not appropriate with LSIs (the use of which on a normally distributed population yields a Galtonian distribution).

A serious problem with the way that some investigators have reported the results of size index analyses is that they have not documented their "standard animals" either with reference to published standards or by listing the standard dimensions they use. Following what was stated earlier about the effects of using different standards to evaluate the same collection of material, it is essential that investigators be able to evaluate the possible effects of choice of standards on the size index distributions. This can be done only if all standards employed are available as a matter of public record. Also, even if the individual dimensions of archaeological specimens are not published, having the standards available permits other investigators to use those same standards to scale their own material for direct comparison with the diagrams in the original article.

Standards do not need to be derived from modern specimens. Thus the Degerbøl cattle standard employed in figures 7 and 8 (and reported in table 1) is calculated from prehistoric material. But that material comprises complete or partially complete skeletons. More problematic is the use of common bone refuse to formulate a standard. One does not know what biases may be reflected in the resulting means, particularly those calculated from small numbers of specimens. If one uses a standard derived from common bone refuse, it is important that large collections of carefully identified specimens be employed, that the nature of the distribution of the values for each dimension be determined (normal?, bimodal?) and that the coefficient of variation be reasonable given the amount of sexual dimorphism demonstrated or expected. In fact, the same guidelines are also appropriate for standards derived from modern specimens, particularly if the VSI technique is to be employed.

Recommendations with respect to the use of size indices

Following from what I have noted above, if size indices are to be employed, I recommend that the researcher:

1) state the formula(s) for the size index technique(s) used;
2) either include a table of standard dimensions employed or refer to a previously published listing of those dimensions;
3) do not combine indices for length and breadth/depth dimensions on the same graph;
4) explicitly define any intervals used to plot the results;
5) calculate and plot the locations of the median (mid-point of the distribution) and the mean and compare the two (to evaluate how skewed the distribution is);
6) evaluate the effects of using different standards and different size index techniques;
7) be aware, if the LSI technique is employed, that the log of the mean of two values does not equal the mean of the logs of the same two values and that an LSI distribution is non-linear;
8) be aware, if the VSI technique is employed, that a third variable, s, is introduced into the calculation of the index (in addition to x and m, which are the only variables used by the SI, RSI, and LSI techniques); this is likely to make the VSI distribution different from that of SI, RSI, or LSI.
9) evaluate the effects of index averaging of measurements taken on the same specimen or articulating specimens (this can be done by plotting the results using different averaging and not-averaging protocols);

10) experiment with different graphic presentations of the material; different plots highlight different aspects of size index distributions;

11) do not over-interpret the results.

Conclusion

Size index techniques were developed to deal with the problems posed by small assemblages of measurable bone. In an attempt to identify patterns of change and variability in animal size, analysts have sought to combine measurement information from different skeletal elements. In so doing, however, element-specific information is lost and, as discussed above, certain simplifying assumptions have to be made. These place limitations on what can legitimately be demonstrated using the techniques.

One of the principal uses of size index techniques has been to investigate size diminution postulated to have accompanied the domestication particularly of bovids and pigs. As some analysts have pointed out, however, a change in kill-off patterns may lead to an impression of overall size diminution of a “population” as reflected in the distribution of bone measurements. For example, Ducos (1991) has argued that size decreases evident in the cattle LSI graphs of Helmer (1989) may be due more to a shift in the sex of the animals providing the measured bones than to overall diminution in body size. A similar argument has been made by Zeder (1998a; 1998b) with respect to the sheep and goat RSI graphs of Uerpmann (1979). She argues that a shift of the distribution toward a smaller size can indicate that the bones of fewer males are included in the diagrams because they were killed off before many of their bones had fused. In addition, Zeder (ibid.) makes the point that body size is affected by environment. Thus using an animal from one region to provide the standard for a more distant and environmentally different region may create a false impression of size diminution. The way to avoid this problem, of course, is to examine trends through time region by region and to compare the trends and not the absolute positions of the values on the size index diagrams. Thus while I used the Uerpmann (1979) standards from Western Asia for evaluating sheep and goat material from Mehrgarh in Pakistan (e.g., Meadow 1981; 1984; 1991; 1993), I was looking for trends over time at Mehrgarh and not investigating the relationship between animal sizes in Pakistan and Western Asia.

Size index distributions, whether they be SI, LSI, RSI, or VSI, must be seen as rather coarse reflections of the size characteristics of a population. Their use should not replace the examination of measurements on an element by element basis. Because size indices are derived values, they are secondary data, subject to the effects of the various assumptions and problems noted above. A much more powerful use of measurement data is that employed, for example, by Hesse (1984) in his elegant study of the Ganj Dareh sheep and goat material. By plotting dimensions for each skeletal part separately without the use of any scaling technique, he was able to demonstrate differences between the kill-off of sheep and goat by sex and age. Hesse’s approach was suggested to him by the publications of Higham (e.g., 1967; 1968), who was aware of the use of measurements by German-speaking colleagues because of his research on Central European materials. In the case of both Hesse and Higham, however, large bodies of measured material were available for analysis.

There is a long tradition of osteometric documentation and analysis in Central Europe. This tradition has served to provide archaeozoologists with an enormous body of metrical data for that region. Slowly a similarly comprehensive corpus is being assembled for the Middle East. This has been stimulated by the work of Joachim Boessneck and Angela von den Driesch, their students, and the students of their students, although the pioneering work of others, including particularly Pierre Ducos (1968), should not be overlooked. An especially important contribution to this process was the publication by Angela von den Driesch (1976) of “A Guide to the Measurement of Animal Bones from Archaeological Sites”. As data were collected from Neolithic sites, attempts were made to look for patterns of size diminution across the region. The development and use of size index techniques was the result (e.g., Ducos 1968; Uerpmann 1979; Grigson 1989). With the accumulation of more and more data, however, it is becoming increasingly evident that there was marked sub-regional variability.
in the timing and course of animal domestication, in the adoption of domestic animals, and in the nature of their exploitation. Size index techniques will continue to play a role in the investigation of these issues, particularly when there are few measurements available from analytical units. However, it is essential that the assumptions behind, and the limitations of, size indices be recognized and that they be employed as only one of a number of techniques useful for investigating animal exploitation patterns in the past.

Acknowledgements

Angela von den Driesch contributed in an essential way to my archaeozoological training, and I am deeply grateful for her help and encouragement particularly during the early part of my career. I greatly value her as a colleague as I also do the other archaeozoologists to whose work I refer in this essay. I also greatly appreciate the stimulation provided by a generation of up and coming scholars among whom are Aaron Stutz and Jill A. Weber who provided valuable comments on a draft of this essay.

Bibliography


Table 1 Summary statistics for *Bos primigenius* measurements published by Degerbøl (1970).

<table>
<thead>
<tr>
<th></th>
<th>Total Assemblage</th>
<th>Males only</th>
<th>Females only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s</td>
<td>v</td>
</tr>
<tr>
<td>Scapula BG</td>
<td>66.9</td>
<td>6.7</td>
<td>10.0</td>
</tr>
<tr>
<td>Humerus Bd</td>
<td>104.7</td>
<td>7.5</td>
<td>7.1</td>
</tr>
<tr>
<td>Radius BFp</td>
<td>100.3</td>
<td>7.9</td>
<td>7.9</td>
</tr>
<tr>
<td>Radius Bd</td>
<td>104.5</td>
<td>9.3</td>
<td>8.9</td>
</tr>
<tr>
<td>Metacarpal Bp</td>
<td>79.3</td>
<td>7.7</td>
<td>9.7</td>
</tr>
<tr>
<td>Metacarpal Bd</td>
<td>78.6</td>
<td>7.3</td>
<td>9.3</td>
</tr>
<tr>
<td>Femur DC</td>
<td>62.1</td>
<td>5.0</td>
<td>8.0</td>
</tr>
<tr>
<td>Tibia Bd</td>
<td>83.4</td>
<td>7.0</td>
<td>8.4</td>
</tr>
<tr>
<td>Talus GLI</td>
<td>87.4</td>
<td>5.2</td>
<td>5.9</td>
</tr>
<tr>
<td>Talus Bd</td>
<td>57.9</td>
<td>5.4</td>
<td>9.3</td>
</tr>
<tr>
<td>Cent+IV Tars GB</td>
<td>76.9</td>
<td>8.6</td>
<td>11.2</td>
</tr>
<tr>
<td>Metatarsal Bp</td>
<td>65.1</td>
<td>6.7</td>
<td>10.3</td>
</tr>
<tr>
<td>Metatarsal Bd</td>
<td>73.7</td>
<td>6.3</td>
<td>8.5</td>
</tr>
<tr>
<td>mean</td>
<td>7.0</td>
<td>8.8</td>
<td></td>
</tr>
</tbody>
</table>

Note: Figures 7 and 8 were formulated using the Total Assemblage as the standard. Measurement abbreviations follow von den Driesch (1976).
Table 2 *Gazella* measurements and summary statistics.

**Gazella bennetti** from Eastern Iran

<table>
<thead>
<tr>
<th>NMNH#</th>
<th>328578</th>
<th>328577</th>
<th>329354</th>
<th>Female</th>
<th>Male</th>
<th>Male mean</th>
<th>s</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scapula BG</td>
<td>17.30</td>
<td>17.80</td>
<td>18.75</td>
<td>17.95</td>
<td>0.74</td>
<td>4.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humerus BT</td>
<td>21.30</td>
<td>21.40</td>
<td>22.90</td>
<td>21.87</td>
<td>0.90</td>
<td>4.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radius Bp</td>
<td>23.40</td>
<td>23.50</td>
<td>25.20</td>
<td>24.03</td>
<td>1.01</td>
<td>4.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radius Bd</td>
<td>21.40</td>
<td>21.00</td>
<td>22.65</td>
<td>21.68</td>
<td>0.86</td>
<td>3.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ulna DPA</td>
<td>18.00</td>
<td>18.90</td>
<td>20.65</td>
<td>19.18</td>
<td>1.35</td>
<td>7.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metacarpal Bp</td>
<td>18.45</td>
<td>19.50</td>
<td>20.00</td>
<td>19.32</td>
<td>0.79</td>
<td>4.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metacarpal Bd</td>
<td>17.25</td>
<td>17.90</td>
<td>19.90</td>
<td>18.35</td>
<td>1.38</td>
<td>7.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pelvis LA</td>
<td>23.50</td>
<td>22.30</td>
<td>25.80</td>
<td>23.87</td>
<td>1.78</td>
<td>7.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Femur DC</td>
<td>16.40</td>
<td>16.10</td>
<td>18.25</td>
<td>16.92</td>
<td>1.16</td>
<td>6.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tibia Bd</td>
<td>21.10</td>
<td>20.20</td>
<td>21.70</td>
<td>21.00</td>
<td>0.75</td>
<td>3.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talus Bd</td>
<td>13.90</td>
<td>14.55</td>
<td>14.80</td>
<td>14.42</td>
<td>0.46</td>
<td>3.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Centroquartal GB</td>
<td>17.85</td>
<td>18.20</td>
<td>19.35</td>
<td>18.47</td>
<td>0.78</td>
<td>4.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metatarsal Bp</td>
<td>16.80</td>
<td>17.40</td>
<td>18.00</td>
<td>17.40</td>
<td>0.60</td>
<td>3.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metatarsal Bd</td>
<td>18.40</td>
<td>18.80</td>
<td>21.55</td>
<td>19.58</td>
<td>1.71</td>
<td>8.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calcaneum GL</td>
<td>50.90</td>
<td>53.65</td>
<td>55.85</td>
<td>53.47</td>
<td>2.48</td>
<td>4.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ph1 GLpe posterior</td>
<td>33.40</td>
<td>36.70</td>
<td>35.40</td>
<td>35.17</td>
<td>1.66</td>
<td>4.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ph1 GLpe anterior</td>
<td>37.30</td>
<td>41.20</td>
<td>39.40</td>
<td>39.30</td>
<td>1.95</td>
<td>4.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ph1 GLpe average</td>
<td>35.30</td>
<td>38.90</td>
<td>37.40</td>
<td>37.20</td>
<td>1.81</td>
<td>4.86</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Gazella gazella** from Israel

<table>
<thead>
<tr>
<th>MCZ#</th>
<th>54414</th>
<th>54415</th>
<th>54417</th>
<th>Female</th>
<th>Male</th>
<th>Male mean</th>
<th>s</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scapula BG</td>
<td>19.20</td>
<td>22.50</td>
<td>19.90</td>
<td>20.53</td>
<td>1.74</td>
<td>8.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humerus BT</td>
<td>22.60</td>
<td>24.10</td>
<td>23.00</td>
<td>23.23</td>
<td>0.78</td>
<td>3.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radius Bp</td>
<td>24.80</td>
<td>26.70</td>
<td>25.00</td>
<td>25.50</td>
<td>1.04</td>
<td>4.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radius Bd</td>
<td>21.10</td>
<td>23.60</td>
<td>23.10</td>
<td>22.60</td>
<td>1.32</td>
<td>5.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ulna DPA</td>
<td>20.20</td>
<td>24.90</td>
<td>22.30</td>
<td>22.47</td>
<td>2.35</td>
<td>10.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metacarpal Bp</td>
<td>19.80</td>
<td>22.00</td>
<td>20.00</td>
<td>20.60</td>
<td>1.22</td>
<td>5.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metacarpal Bd</td>
<td>20.00</td>
<td>21.60</td>
<td>20.40</td>
<td>20.67</td>
<td>0.83</td>
<td>4.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pelvis LA</td>
<td>23.70</td>
<td>27.60</td>
<td>25.80</td>
<td>25.70</td>
<td>1.95</td>
<td>7.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Femur DC</td>
<td>17.00</td>
<td>19.00</td>
<td>18.00</td>
<td>18.00</td>
<td>1.00</td>
<td>5.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tibia Bd</td>
<td>22.60</td>
<td>24.80</td>
<td>22.50</td>
<td>23.30</td>
<td>1.30</td>
<td>5.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talus Bd</td>
<td>15.90</td>
<td>17.40</td>
<td>16.50</td>
<td>16.60</td>
<td>0.75</td>
<td>4.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Centroquartal GB</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metatarsal Bp</td>
<td>19.00</td>
<td>21.60</td>
<td>19.70</td>
<td>20.10</td>
<td>1.35</td>
<td>6.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metatarsal Bd</td>
<td>21.80</td>
<td>23.30</td>
<td>21.90</td>
<td>22.33</td>
<td>0.84</td>
<td>3.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talus GLI</td>
<td>26.00</td>
<td>30.10</td>
<td>28.30</td>
<td>28.13</td>
<td>2.06</td>
<td>7.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calcaneum GL</td>
<td>54.70</td>
<td>60.70</td>
<td>58.10</td>
<td>57.83</td>
<td>3.01</td>
<td>5.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ph1 GLpe post.</td>
<td>38.40</td>
<td>40.40</td>
<td>40.10</td>
<td>39.63</td>
<td>1.08</td>
<td>2.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ph1 GLpe ant.</td>
<td>40.60</td>
<td>43.40</td>
<td>43.00</td>
<td>42.33</td>
<td>1.51</td>
<td>3.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ph1 GLpe av.</td>
<td>39.50</td>
<td>41.90</td>
<td>41.60</td>
<td>41.00</td>
<td>1.31</td>
<td>3.19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The *Gazella bennetti* specimens are housed in the National Museum of Natural History, Smithsonian Institution, Washington, DC. The *Gazella gazella* specimens are housed in the Museum of Comparative Zoology, Harvard University, Cambridge, MA. Measurement abbreviations follow von den Driesch (1976).