AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 90:377–379 (1993)

Notes and Comments

Human Cranial Variability: A Methodological Comment

Mike Foote

Museum of Paleontology, University of Michigan, Ann Arbor, Michigan 48109-1079

Measurement of morphological variation is essential to our understanding of human evolution. At least two classes of statistics can be used to describe variation: (1) statistics such as the variance that measure *average* dissimilarity (distance), and (2) statistics such as the range of variation that measure *maximal* dissimilarity. Either kind of measure may be meaningfully employed depending on one's interest, but each has radically different sampling properties that must be taken into account. Sampling properties are particularly important because we generally want to infer, from the sample we happen to have, aspects of the underlying distribution from which the sample was drawn. In a recent letter, Wolpoff (1992) compares cranial variation in a sample of 13 Paleolithic humans from the Levant to that in a sample of 388 modern humans from London. By basing his conclusions on the sample range and failing to consider sampling bias associated with this statistic, Wolpoff seriously misrepresents the difference in variability between Paleolithic and modern humans.

It has long been known that, in addition to being very sensitive to outliers, the range depends strongly on sample size (Pearson, 1926). This obvious dependence partly underlies the preferred use of sample statistics such as the variance (which reflects *average* difference), and related statistics, as measures of dispersion (Van Valen, 1974; Uytterschaut and Wilmink, 1983; Franciscus and Long, 1991; van Vark and Bilsborough, 1991). The sample variance provides an unbiased estimate of the parametric variance (Fig. 1A) (Sokal and Rohlf, 1981). However, the sample range is generally biased, increasing monotonically with sample size. Because a biological population is of finite size, it does have a true range that is knowable in principle, but one must sample the entire population to determine this range. The smaller a sample one draws from a population, the smaller the sample range will be. This point is particularly important if we compare large modern samples to small fossil samples.

Dependence between sample size and range of variation can be conveniently illustrated with a normal distribution (Fig. 1B). Although the nature of this dependence is determined here by simulation, it can also be determined analytically (Pearson, 1926). Given a univariate normal distribution, regardless of its variance, expected sample range increases roughly as the logarithm of sample size (Rohlf and Sokal, 1981; see also Foote, 1992). In a comparison between $N = 13$ Paleolithic humans and $N = 388$ modern humans, we would expect the range of the modern sample to be about 1.8 times that of the Paleolithic sample if the two samples were drawn from normal distributions with the same variance (Fig. 1C). This expected difference is remarkably similar to the twofold difference in range that Wolpoff finds for some cranial measures! The apparent difference in variation between the two samples probably reflects sampling bias much more than it shows true differences in variation between the distributions that these samples represent. Wolfpoff's claim that "the *amount* of variation in measurements from the Middle Paleolithic . . . appears to be less than that in a modern population" (p. 142; emphasis in original), is unjustified, unless (1) his emphasis is on the *appearance* rather than the existence of a true difference in variation, or (2) his emphasis is only on the samples, rather than the underlying populations that these sam-

Received May 4, 1992; accepted June 28, 1992.

ples represent. These two possibilities do not reflect the most interesting evolutionary questions.

Wolpoff also suggests that simply discarding the largest and smallest extremes from the larger sample should correct for sampling bias, but this is clearly untrue (Fig. 1C). Even if this correction is made, samples of $N = 388$ and $N = 13$ from a single normal distribution are expected to differ by a factor of 1.6 in their ranges of variation (Fig. 1C). Comparison of ranges, with or without Wolpoff's correction, provides no evidence for Wolpoff's claim that modern populations are more variable than their Paleolithic counterparts.

I do not claim that the samples in question were drawn from normal (much less multivariate normal) distributions. I have chosen to illustrate sample-size effects with a normal distribution, but the existence of sample-size effects is a general property of morphological distributions. The exact relationship between range of variation and sample size depends on the details of this distribution. Sample-size dependence for distributions that are too complex to be easily characterized can be studied by random resampling at different sample sizes (rarefaction), thereby correcting for sample-size differences (Foote, 1992). Other methods of resampling and randomization are very useful when comparing populations with small and/or unequal sample sizes and complex or unknown statistical distributions (Efron, 1982). Regardless of the details of a distribution, if a measure of variation depends on sample size—as the range does—then only

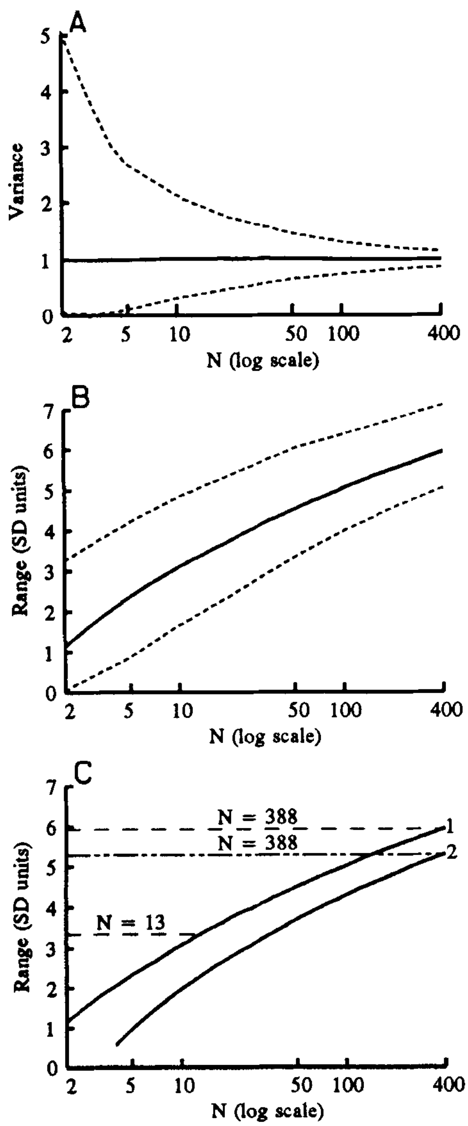

Fig. 1. Relationship between sample size (N) and statistics of variation. Solid curves give expected statistic at given sample size; dashed curves give 95% confidence limits about the expectation. Statistics are determined for each sample size from 2 to 400; each point based on 1,000 random samples drawn from simulated, standard normal distribution ($\mu = 0$, $\sigma = 1$). Similar relationships hold for other distributions (Foote, 1992). Breadth of confidence limits is an indication of uncertainty in sample statistic. Ranges are in standard deviation (SD) units and therefore hold for any normal distribution. **A.** Sample variance provides unbiased estimate of parametric variance. Uncertainty is high at small sample sizes and decreases relatively rapidly with increasing sample size. **B.** Sample range increases monotonically with sample size, and uncertainty decreases relatively slowly with increasing sample size. **C.** Comparison of ranges of samples with different N drawn from same distribution. *1*, samples from standard normal distribution; *2*, samples from standard normal distribution, highest and lowest extremes discarded, according to Wolpoff's (1992) correction. Expected range for sample sizes of $N = 13$ (curve 1 only) and $N = 388$ (curves 1 and 2) are indicated. Samples of $N = 388$ are expected to have range 1.8 and 1.6 times as large as samples of $N = 13$ (for curves 1 and 2, respectively). Thus ranges of *samples* may differ appreciably, but this does not imply difference in variation in the underlying distribution from which samples are drawn. Note that curve in B is the same as curve 1 in C.

samples of the same size should be compared. Whatever Wolpoff's substantive claims about human variation, his approach is fundamentally misleading and should be categorically avoided by anthropologists addressing the question of relative levels of variation in different populations.

ACKNOWLEDGMENTS

I thank P.D. Gingerich, D.W. McShea, B.H. Smith, M.H. Wolpoff, and an associate editor for comments and discussion. Acknowledgment is made to the Donors of the Petroleum Research Fund, administered by the American Chemical Society, for support of this research.

LITERATURE CITED

Efron B (1982) The Jackknife, the Bootstrap, and Other Resampling Plans. Philadelphia: Society for Industrial and Applied Mathematics.

Foote M (1992) Rarefaction analysis of morphological and taxonomic diversity. Paleobiology 18:1–16.

Franciscus RG, and Long JC (1991) Variation in human nasal height and breadth. Am. J. Phys. Anthropol. *85:*419–427.

Pearson ES (1926) Further note on the distribution of range in samples taken from a normal population. Biometrika *18:*173–194.

Rohlf FJ, and Sokal RR (1981) Statistical Tables, 2nd ed. San Francisco: Freeman.

Sokal RR, and Rohlf FJ (1981) Biometry, 2nd ed. San Francisco: Freeman.

Uytterschaut HT, and Wilmink FW (1983) On the assumption of equality of variance-covariance matrices in the sex and racial diagnosis of human skulls. Am. J. Phys. Anthropol. *60:*347–357.

Van Valen, L (1974) Multivariate structural statistics in natural history. J. Theoret. Biol. *45:*235–247.

van Vark GN, and Bilsborough A (1991) Shaking the family tree. Science *253:*834.

Wolpoff MH (1992) Levantines and Londonders. Science *255:*142.