Statistical inference, scale and noise in comparative anthropology

To the Editor — In an insightful Comment Bliege Bird and Codding[1] highlight a number of important issues to consider in the analysis of cross-cultural anthropological data. However, a casual reader of the Comment could be forgiven for taking away the message that cross-cultural data in anthropology is inherently flawed, and so is of limited use. We want to emphasize that comparative analysis plays an essential role in all non-experimental sciences, including anthropology and archaeology. This is because when systems cannot be manipulated due to scales of time and space, or issues of logistics or ethics, the only way to evaluate alternative outcomes is by analysing the results of natural experiments.

Human societies are complex, adaptive, noisy, scale-dependent, hierarchical, self-organizing, non-ergodic systems, exhibiting emergent statistical features at all scales. It is simply not possible to understand the structure and dynamics of a complex system by observing a single scale, no matter how well studied that scale may be, thus we must combine top-down inference with bottom-up observation. Similarly, it is not possible to understand large-scale patterns by studying single case studies, no matter how well-studied those case studies may be. Furthermore, comparative analysis is central to interdisciplinary science, revealing deep theoretical insights between traditionally disparate areas of research, hence the explosive success of complex systems science in recent years[2,3].

Bliege Bird and Codding[1] raise the important issue of data quality in comparative research and point to particular examples of the Northern Aranda and the Martu. We argue these examples highlight the importance of statistical inference in comparative anthropology. All comparative data sets in all disciplines are unavoidably noisy as data points are estimates of unobserved phenomena, synthesized from the work of multiple researchers often working in very different circumstances for very different reasons. In this sense, all data are subject to false precision. However, noisy data are not necessarily biased data, and having multiple estimates of single data points is actually desirable. Consider the case of the Aranda. If we wish to estimate pre-contact population size but have no prior demographic estimate, we must sample the literature, which yields two estimates seven years apart: 2,000 in 1920 and 300–400 in 1927[1]. The fact these two estimates differ is not a problem, and is in fact useful. We have just updated our inference from no information to two bits of information, which, by definition, measurably reduced the uncertainty (that is, the entropy) of our estimate, and we can now bound our estimate of Aranda pre-contact population size. Maximum likelihood would suggest a principled move from here is to take the average of the two estimates. What we shouldn't do is disregard the Aranda information we have just acquired unless there is a statistically principled reason to do so, such as blatant bias, error, or as a statistically significant outlier with undue leverage. To ignore the estimates otherwise would be to introduce additional uncontrolled observer bias into the data. Statistically, we should never arbitrarily reduce sample sizes as we are in the job of trying to explain as much variation as possible by accounting for as much error as possible. Statistical power comes from sample size not precision.

Perhaps more importantly, comparative datasets generate statistical distributions that provide perfectly reasonable inferences of parameters, no matter how noisy the data. This is because the central limit theorem and the law of large numbers tells us that the statistical issue is not how precise each observation may be, but whether a large sample of similarly noisy observations provides an unbiased, well-behaved estimate of a parameter, such as an average, a variance or a scaling exponent. Noise simply affects the error bars around the parameter estimate, not the value of the parameter itself. That is to say, the power of comparative data analysis is not the reliability (or otherwise) of any individual data point, but the statistical inference of parameters from distributions of data that are unobservable at smaller scales.

We need to remember that noise is not unique to comparative datasets, nor is it unique to anthropological datasets. All data collected in all fields at all scales are measured with error; the statistical challenge is whether we can control that error to find the underlying signal we are interested in. This is not to say all data are good data; of course, biased data are bad data no matter what we do to them. However, noisy data are not necessarily biased data, and noise is not on its own a signal of bad data. As Bliege Bird and Codding[1] emphasize, statistical issues of autocorrelation and non-independence in cross-cultural data are commonplace, but can be addressed. Similarly, we argue, noisy data should not be avoided, but addressed and embraced. ❐

Marcus J. Hamilton[1]✉ and Miikka Tallavaara[2]

[1]Department of Anthropology, University of Texas at San Antonio, San Antonio, TX, USA. [2]Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland.

✉*e-mail: marcus.hamilton@utsa.edu*

Published online: 16 December 2021
https://doi.org/10.1038/s41559-021-01637-3

References

1. Bliege Bird, R. & Codding, B. F. *Nat. Ecol. Evol.* https://doi.org/10.1038/10.1038/s41559-021-01579-w (2021).
2. West, G. *Scale: The Universal Laws of Growth, Innovation, Sustainability, and the Pace of Life in Organisms, Cities, Economies, and Companies* (Penguin, 2017).
3. Bettencourt, L. M. A. *Introduction to Urban Science: Evidence and Theory of Cities as Complex Systems* (MIT Press, 2021).

Competing interests

The authors declare no competing interests.