Professional Zooarchaeology Group (PZG) Minutes

Statistics, University of Nottingham, 17th January 2009

On 17th January 2008 the eighth PZG meeting was held at the Department of Archaeology University of Nottingham. The meeting was organized by Polydora Baker and Naomi Sykes on the subject of Statistical Analysis in Zooarchaeology. The day was well attended.

The excellent presentations expounding and explaining the use of statistics for zooarchaeological data analysis were given by Hannah O'Regan (Liverpool John Moores University) and Andrew Millard (University of Durham), and examples were predominantly based on the data from the LIA/Roman site Elms Farm, Heybridge in Essex (Johnstone and Albarella, 2002; Albarella *et al.*, 2008). Datasets from this site were also circulated prior to the meeting to allow audience participation in the analyses demonstrated. Extensive notes on statistics in zooarchaeology were provided by Jen Harland (Fishlab, University of York)

All of the analyses demonstrated during the workshop used the PAST statistical package, which provides a comprehensive range of easy to use analytical statistical software and can be downloaded for free from http://folk.uio.no/ohammer/past/download.html.

The initial presentation of the day was given by Hannah, in which she introduced the key concepts involved in statistical analysis, explaining the rationale and terminology commonly used in statistical manipulation. After explaining the aims of statistical analysis, she went on to define concepts such as samples, variables and objects and explained the different types of data used in statistical analysis (continuous, discontinuous, nominal, ordinal, interval and ratio). She then focused on descriptive statistics, and summary statistical calculations such as mean, mode and median were explained along with range, variance, standard deviation, coefficient of variance, and standard error. The relationship of these statistics to the distribution of data (Normal, bimodal etc) was expounded. Hannah then explained the importance of identifying kurtosis in data before choosing the appropriate statistical techniques to use. She introduced us to the use of Chi Squared, Shapiro-Wilk and Jarque-Bera tests in order to check for normality.

Following a coffee break Andrew took over and introduced the basic concepts of statistical testing. Having explained the null hypothesis, p-values and probabilities, he went on to demonstrate simple parametric statistical tests such as the t-test using PAST, an activity which the group joined in with. Having achieved our result Andrew cautioned the group about possible misinterpretations of the P-value returned, and explained about type 1 and type 2 errors, and how to identify and avoid these, advising careful consideration of the assumptions used in the test. Andrew then discussed choosing the appropriate statistical test to use according to the questions asked of the data and the nature of the data itself. After explaining the difference between parametric and non-parametric tests, he reviewed the characteristics and limitations of different comparative statistical techniques – using counts: Chi-squared, Fishers exact test, bimodal test; and averages: t-Test, Mann-Whitney U test, before moving on to explain multivariate comparisons of average: ANOVA and the Kruskal-Wallis test, and comparisons of distributions such as the Kolmogorov-Smirnov test and Shapiro-Wilk test. Following this the group were invited to consider the appropriate test to use in a variety of analytical situations, and to practice using Chi-Squared tests in PAST.

After a buffet lunch, Hannah took to the podium again and explained the application and limitations of a number of parametric tests including the Mann-Whitney U Test for comparing two samples, and the Kruskal-Wallis test for multiple samples. This was followed by comparisons of distributions between two samples using a Kolmogorov-Smirnov test, and for multiple samples

using the Freidman test. For most of the methods explained the group were asked to analyse sample datasets from the Elms Farm biometrical data using PAST.

Andrew then took over and after quickly reviewing chi squared tests and Fishers exact test, he introduced the group to multivariate statistics in the form of correspondence analysis. We were invited to use data on fish exploitation provided by Jen Harland (see Barrett et al. 2008; Barrett et al. 2004a; Barrett et al. 2004b) to produce CA plots in PAST, which were then interpreted in terms of temporal patterns in species distribution.

Hannah then briefly introduced diversity indices using Simpson's index, followed by cluster analysis, before the group broke for coffee.

After the coffee break, during which the subject and location of the ninth PZG meeting were discussed, and the organisers and speakers of the current meeting thanked, time was provided for PZG members to work on the analysis of their own data in the light of the day's presentations, with assistance from Andrew and Hannah.

Our thanks go to the organisers Polydora Baker and Naomi Sykes, to the speakers, Hannah O'Regan and Andrew Millard, and to contributor Jen Harland for putting together an extensive, informative and useful introduction to statistical analyses in zooarchaeology.

Over the course of the day the following resources were recommended as useful:

Books:

Shennan, S. 1997. *Quantifying Archaeology*, Edinburgh University Press.

Fowler, J. Cohen, L. and Jarvis, P. 1998. *Practical statistics for field biology*. Wiley, Chichester

Dytham, C. 2003. *Choosing and using statistics: a biologists guide*, Blackwell, Oxford.

Web resources:

Deacon, J (n.d.) The really easy statistics site, University of Edinburgh: http://www.biology.ed.ac.uk/research/groups/jdeacon/statistics/tress1.html

Dataset for digging numbers – Quantitative archaeology wiki. (to be used in conjunction with: Fletcher, M and Lock, G.R. 2005. *Digging Numbers*. Oxford University Committee for Archaeology monograph): http://wiki.iosa.it/diggingnumbers:start

R- package for statistics used by professional statisticians.- programming involved. http://www.r-project.org/

SOCR: Statistics Online Computational Resource: http://www.socr.ucla.edu/

Data sets were used from:

Johnstone, C and Albarella, U. 2002. *The Late Iron Age and Romano-British Mammal and Bird Bone Assemblage from Elms Farm, Heybridge, Essex (Site Code: HYEF93e95)*. Centre for Archaeology Report 45/2002. English Heritage, Portsmouth.

See also:

Albarella, U., Johnstone, C. and Vickers, K. 2008. The development of husbandry systems from the Late Iron Age to the end of the Roman period: a case study from South-East Britain. *Journal of Archaeological Science* **35**, 1828-48.

Minutes contributed by Kim Vickers

If you would like this document in a different format, please contact our Customer Services department:
Telephone: 0870 333 1181
Fax: 01793 414926
Textphone: 01793 414878
E-mail: customers@english-heritage.org.uk