Acta Theriologica 37 (1 – 2): 199 – 203, 1992.
PL ISSN 0001 – 7051

COMMENTS

An incorrect analysis of archeozoological data on the aurochs

Jerzy BAŃBURA

Bańbura J. 1992. An incorrect analysis of archeozoological data on the aurochs. Acta theriol. 37: 199 – 203.

Serious errors in the statistical analysis conducted in a paper on the aurochs *Bos primigenius* (Lasota-Moskalewska and Kobryń 1990) are revealed in this article. A linear coding method transforming raw measurements into dimensionless numbers ranging from 0 to 100 was applied by the authors. Pooling the coded data on different traits is criticised and so is their application to any statistical tests. Correct tests are proposed.

Department of Ecology and Vertebrate Zoology, University of Łódź, Banacha 12/16, 90-237 Łódź, Poland

Key words: auroch bones, statistical tests, mistakes, coding data, inference

Introduction

In a recent paper Lasota-Moskalewska and Kobryń (1990) analysed osteometric data on the aurochs in order to investigate whether or not bone dimensions of this species underlie any temporal and/or spatial trends. I believe that this paper is an example of unacceptable misuse of statistical methods, perhaps, the most incorrect data treatment I have ever seen.

It is well known that statistical tests serve inferring properties of big collections of things (populations) on the basis of properties of small subsets of these collections (samples). To achieve this aim the fundamental assumption of random sampling is made, which ensures that the sample represents the population. It means eventually that individual measurements should be independently, with no *a priori* relations between each other, taken from the probability distribution of traits in the population. This assumption is even more essential than that of a particular kind of distribution (normal or non-normal). It is possible to draw every conclusion when data do not fulfil the random sampling and independence assumption (Hull 1973, James and McCulloch 1985).

Another possible source of inference errors is the application of inappropriate tests to studied data sets (Chatfield 1988, Gilbert 1989). It is also worth pointing out that even results of a well designed statistical analysis may be misunderstood or cannot be understood at all if their way of presentation is inadequate.

I shall now present what I consider erroneous in Lasota-Moskalewska and Kobryń's (1990) data handling, testing and presentation. To be more readable I shall illustrate some points with the data on three metric features of one-age-group males of *Microtus oeconomus* from Norway, analysed in detail by Markowski and Østbye (1992).

Data treatment

Material for the paper in question consists of archeozoological remains and fragments whose identity (belonging to particular specimens) could not usually be established. Twenty five metric traits of 12 different bones were measured, up to 4 measurements being taken from some bones. Although the authors do not state it explicitly, it is evident that a number of measurements concern different traits of the same bones of the same aurochs individuals, which can be adequately controlled given an appropriate procedure, and that another set of measurements concerns different bones whose identity in the above sense is unknown, and, consequently, cannot be controlled by any statistical method. The raw data are presented using descriptive statistics such as: mean, standard deviation and range, but are not subjected to any tests.

Instead, the authors analyse measurements expressed on what they term a point scale and which in fact means a coding procedure employing the formula below

$$Y_i = [(X_i - X_{min}) / (X_{max} - X_{min})]\ 100\%, \quad (1)$$

where: Y_i – a coded measurement of a given trait of an i-th individual; X_i – a raw measurement of a given trait of an i-th individual; X_{max} and X_{min} – maximum and minimum raw measurement values of a given trait recorded in the whole data set.

It is easy to see that this coding procedure uses a linear transformation

$$Y_i = bX_i + a, \quad (2)$$

as (1) can be expressed, omitting per cent, as follows

$$Y_i = X_i\ [1/(X_{max} - X_{min})] - [X_{min}/(X_{max} - X_{min})] \quad (3).$$

Having their data coded, the authors treat them as "comparable irrespective of type and size of the characteristic" (Lasota-Moskalewska and Kobryń 1990, p. 91). As a consequence, after coding different traits are pooled in a number of configurations, combining all characters or some subsets of them as, for instance, 7 characteristics of the forelimb.

Let us consider the properties of the linear coding. Such a procedure affects the raw data average by multiplying it by a constant b and addition of a constant a, but the effect on variance is only due to multiplying by the constant b (Zar

1984). The linear coding neither changes the shape of frequency distributions of original data nor influences their correlations. The latter is particularly worth pointing out, since it evidently shows that the linear coding does not destroy correlations which are usually present among different metric features of organisms. As a consequence, the assumption of independence of individual measurements is violated and using any statistical inference from such a mixed data set is not justified. As a matter of fact, abusage of statistics is also the case when individual traits pooled are not correlated because zero correlation and independence are not equivalent (Broffitt 1986).

Even a brief inspection of Tables 3 and 15 in Lasota-Moskalewska and Kobryń (1990) seems to suggest that the frequency distributions of coded measurements are polymodal compositions, forming bizarre shapes (*sensu* Bradley 1977).

Table 1. Matrix of correlations among three metric traits of the skull of *Microtus oeconomus*. Correlations among raw measurements are displayed above the diagonal and below it are shown correlations between lineary coded measurements. N = 48, all values are significant at $p < 0.01$. Symbols: CbL – condylobasal length; BCL – brain case length; DL – diastema length.

Trait	CbL	BCL	DL
CbL	–	0.86	0.76
BCL	0.86	–	0.57
DL	0.76	0.57	–

Three metric traits of the skull of *Microtus oeconomus* may be an empirical example of what was said about the properties of linear coding with respect to correlation. Table 1 shows that the coefficients of correlation among the three traits have exactly the same values after coding as before it. It would be a violation of the independence assumption if the coded values of these three traits were pooled and used to test statistical hypotheses because they would not form 144 independent data points (as assumed in the criticised paper) but just 48 clusters of intrcorrelated points. By analogy, the same could be shown using Lasota-Moskalewska and Kobryń's data on different characteristics of particular bones of the aurochs.

Tests

As it was stated above, any statistical tests are inapplicable to the pooled data set of Lasota-Moskalewska and Kobryń's (1990). The authors, however, compare two time samples and several geographically different samples of bone measurements. In their method section they describe the chi-square test used to this comparison, but as far as I can guess it is only employed to test frequencies of males and females in individual samples (Tables 13 and 25). In other tables presenting results of statistical inference (Tables 4 – 12, 16 – 24 and 26) the Góralski test (Góralski 1966) is used in order to test whether or not empirical frequencies in individual cells of relevant contingency tables deviate from expected frequencies.

Given an appropriate data set, this test is a very useful and original tool being in fact a form of residual analysis (Haberman 1973, Everitt 1977). It is a great pity that the paper describing this test (Góralski 1966) was published in a journal inaccessible to most of potential users. Nonetheless, the results of this test should be interpreted with caution, since the residuals are not independent in contingency tables (Siegel and Costellan, Jr. 1988). As a consequence, it ought to be applied only to the tables with significant chi-square values and to planned comparisons. The more common practice in using chi-square is subdividing contingency tables on the basis of the additivity properties but this of course does not concern single cells (Everitt 1977).

Taking into account the fact that the data analysed in the paper in question are pooled mixtures, it is difficult to figure out what null hypotheses are being tested. The null hypothesis that the bone dimensions are not differentiated temporally and geographically, suggested by the authors, is not fit to their data set. The reason is that an influence of the number of individual measurements of particular traits contributing to different ranges on the point scale on deviations from expected frequencies in contingency tables is not controlled.

Two courses of action can be recommended to do a correct analysis of these aurochs data. Firstly, hypotheses about sexual, temporal and geographical differentiation of particular characteristics presented in Tables 2 and 14 can be directly tested using the Student t-test and analysis of variance, respectively, in all cases where data are abundant enough. Non-parametric methods may be considered as alternatives. Secondly, when measurements of several different traits of individual bones are available, multivariate methods, like principal components analysis, should be applied (Manly 1986).

Presentation

The way of presentation by Lasota-Moskalewska and Kobryń (1990) of both their statistical methods and results may lead to misunderstanding. There is no reason to go into details of the chi-square test that is a common standard method, but it would certainly be desirable to describe an unusual method like the Ψ test step by step.

Titles of tables presenting results of the Ψ test are misleading as they suggest that the tables contain correlations of some sort. Actually, they display empirical and expected frequencies in individual cells of various two-way contingency tables.

Significance of differences between these frequencies is also shown, but there lacks information about overall chi-square values or about Ψ values themselves.

Conclusion

Lasota-Moskalewska and Kobryń (1990) made serious errors in data handling, testing and presentation. First of all, the method of transformation of original data into points is misleading and inapplicable to the analysed data set. Morever, inadequate tests and wrong presentation of results make reading this paper a difficult job.

In general, I cannot help concluding that it is almost impossible for a researcher without a very special background to be capable of inventing a correct new statistical method. Therefore consulting good handbooks, like Sokal and Rohlf (1981) or Zar (1984), and professional statisticians is advised in order to avoid wasting time on ill-designed data analysis that must in consequence give unreliable conclusions.

Acknowledgements: I am very grateful to J. Markowski, Z. Wojciechowski and P. Zieliński for their critical comments on the first draft of this article. I thank J. Markowski for providing me with the vole data.

References

Bradley J. V. 1977. A common situation conducive to bizarre distribution shapes. Amer. Stat. 31: 147 – 150.

Broffitt J. D. 1986. Zero correlation, independence, and normality. Amer. Stat. 40: 276 – 277.

Chatfield C. 1988. Problem solving. A statistician's guide. Chapman and Hall, London: 1 – 261.

Everitt B. S. 1977. The analysis of contingency tables. Chapman and Hall, London: 1 – 128.

Gilbert N. 1989. Biometrical interpretation. Second edition. Oxford Univ. Press, Oxford: 1 – 146.

Góralski A. 1966. Kryterium oceny istotności nadwyżki i niedoboru liczebności w elementarnej kostce wielodzielczej *m*-wymiarowej tablicy. Zeszyty Naukowe Politechniki Warszawskiej 131, Mat. 7: 7 – 32.

Haberman S. J. 1973. The analysis of residuals in cross-classified tables. Biometrics 29: 205 – 220.

Hull D. 1973. How to lie with statistics. Penguin Books, London: 1 – 124.

James F. C. and McCulloch C. E. 1985. Data analysis and the design of experiments in ornithology. [In: Current Ornithology. Vol. 2. R. F. Johnston, ed.]. Plenum Publishing Corp., New York: 1 – 63.

Lasota-Moskalewska A. and Kobryń H. 1990. The size of aurochs skeletons from Europe and Asia in the period from the Neolithic to the Middle Ages. Acta theriol. 35: 89 – 109.

Manly B. F. J. 1986. Multivariate statistical methods–A primer. Chapman and Hall, London: 1 – 159.

Markowski J. and Østbye E. 1992. Morphological variability of a root vole population in high mountain habitats, Hardangervidda, South Norway. Acta theriol. 37: 117 – 139.

Siegel S. and Castellan, Jr. N. J. 1988. Nonparametric statistics for behavioral sciences. Second edition. McGraw-Hill Book Company, New York: 1 – 399.

Sokal R. R. and Rohlf F. J. 1981. Biometry. Second edition. Freeman, San Francisco: 1 – 859.

Zar J. R. 1984. Introduction to statistical analysis. Second edition. Prentice Hall, Inc., Englewood Cliffs: 1 – 718.

Received 16 April 1992.