Animal Behaviour 93 (2014) 183–189

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Commentary

Effective use of Pearson's product–moment correlation coefficient

Marie-Therese Puth [a], Markus Neuhäuser [a], Graeme D. Ruxton [b, *]

[a] Department of Mathematics and Technology, RheinAhrCampus, Koblenz University of Applied Sciences, Remagen, Germany
[b] School of Biology, University of St Andrews, St Andrews, U.K.

ARTICLE INFO

Article history:
Received 14 February 2014
Initial acceptance 18 March 2014
Final acceptance 6 May 2014
Published online
MS. number: 14-00137R

Keywords:
association
confidence internal
null hypothesis testing
regression
Spearman

The calculation of correlation coefficients is widespread in biological research. Often, the null hypothesis of zero correlation is tested and/or confidence intervals for the correlation are computed. There are several different methods for this purpose; we compare the performance of different methods. According to our results the standard *t* test approach does offer generally reasonable performance even when the underlying distribution departs from bivariate normality. However, for non-normal data alternative techniques, especially the permutation test and using the RIN (rank-based inverse normal) transformation, offer better control of type I error and good power. With regard to confidence intervals, all investigated methods perform similarly; and there is no consistent pattern with which to strongly recommend one method over another. However, we show that two easy-to-calculate methods based on asymptotic results often perform tolerably well even for small sample sizes.

It is common in statistical analysis to explore and summarize the strength of association between two interval scale, measured traits on a number of experimental units. In 2013, *Animal Behaviour* published 315 papers involving statistical analysis of empirical data; of these we found that 106 (34%) reported measures of association for interval scale variables. Our aim is to offer advice on how such investigations can most effectively be carried out and presented. Our survey of 2013 *Animal Behaviour* papers suggests that current practice is mixed. For example, of these 106 papers 71 used Pearson's product–moment correlation coefficients to describe the strengths of associations between variables, 25 used Spearman's coefficient and 10 papers used both. None of the papers that exclusively used one type of coefficient stated a reason behind the authors' choice. Of those that used examples of both methods, most stated that prior to association being measured histograms of the two variables were inspected and Spearman's coefficient was used if the distribution of either variable deviated substantially from normal; otherwise Pearson was preferred. A recent publication has surveyed textbooks and found considerable variation in their advice on the assumptions underlying Pearson's coefficient and its robustness to deviations from underlying assumptions (Bishara & Hittner, 2012); hence, we begin our paper with a discussion of these assumptions.

In our survey of 106 papers, 86 papers not only calculated a correlation coefficient from a sample but also tested the null hypothesis that the population value of that coefficient was zero. Of those 86 papers, sufficient information was given to infer the method used for only 32 papers. Hence, we explore the different methods available for testing this hypothesis and offer advice on the most effective in different circumstances. No paper formally tested against a predicted value other than zero, although 42 of the papers informally compared observed estimates with a theoretically predicted value other than zero. Hence, we present a description of how this null hypothesis could be formally tested.

Only three of the 106 papers (3%) offered a confidence interval around any calculated correlation coefficient, despite methods for calculating these being readily available. Of these three papers, only one stated the method used to calculate the confidence interval and in that case the method was not specified in sufficient detail to allow recreation of the calculation. Hence, we finish our paper by offering advice on the most effective methods of calculating such confidence intervals.

PEARSON'S PRODUCT–MOMENT CORRELATION COEFFICIENT (*R*) DEFINED

The most commonly used measure of association is Pearson's product–moment correlation coefficient, often denoted *r*, see e.g.

* Correspondence: G. D. Ruxton, School of Biology, University of St Andrews, St Andrews, Fife KY16 9TH, U.K.
E-mail address: gr41@st-andrews.ac.uk (G. D. Ruxton).

http://dx.doi.org/10.1016/j.anbehav.2014.05.003
0003-3472/

Whitlock and Schluter (2009). If we measure two quantities X and Y on each of N individuals to give a data set $(X_1,Y_1),\ldots,(X_N,Y_N)$, then

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i,$$

$$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i,$$

$$r = \frac{\sum_{i=1}^{N}\{(X_i-\overline{X})(Y_i-\overline{Y})\}}{\sqrt{\sum_{i=1}^{N}(X_i-\overline{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i-\overline{Y})^2}}.$$

WHEN IS A MEASURE OF CORRELATION OR ASSOCIATION A USEFUL STATISTIC?

Assume we have two continuous traits, X and Y, measured on each of the N individuals in a sample. Correlation and simple linear regression are two ways of exploring a potential linear relationship between the values of the two traits; both describe features of a scatterplot. These two methods are often used interchangeably (Whitlock & Schluter, 2009). However, the key to whether regression analysis should be used to study the relation between X and Y depends on our understanding of the underlying biology. If there is functional dependence then regression can be applied. The key to functional dependence is that there is an asymmetry between the variables, such that it makes biological sense to explore whether Y is a function of (i.e. is dependent upon) X, but the converse does not make sense. For example, it would seem rational to measure the blood pressure and record the age of a number of humans, then ask the question 'is blood pressure dependent upon age?', but it would seem strange to ask whether age depends on blood pressure. Here there is functional dependence and a regression can be applied for prediction from one variable to another. There need not be an identified cause-and-effect mechanism underlying functional dependence. Consider the case where we take a sample of primate species from zoo records to obtain a characteristic mass and longevity for each species. Again there is the asymmetry: the question 'is longevity affected by mass across the primates' make more biological sense than 'is mass affected by longevity across the primates'. We need not be certain that we can identify a mechanism by which mass influences longevity; it is sufficient that we think that the effect of mass on longevity is likely to be more direct than the effect of longevity on mass. Finally, if we have conducted an experiment in which we controlled the value of X and varied this between treatment groups and measured Y in each experimental unit, then again we have imposed an asymmetry in our experimental design, and again regression would be appropriate for exploring linear relationships between X and Y.

In situations in which there seems to be symmetry, and it is just as valid to ask how a change in X would be expected to influence Y as vice versa, correlation analysis can be used to explore this association. Examples of such traits might be wing length and tail length across different species of bird, students' scores in mathematics and physics exams, the numbers of flower species and numbers of beetle species found in each of a sample of islands, or the ratio between the lengths of the second and fourth fingers in a sample of humans and a measure of willingness to take risks in those individuals.

Both regression and correlation coefficient approaches offer similar behaviour in terms of testing the null hypothesis of no association. However, the computation of a correlation coefficient is useful to measure the relationship, or association, between variables whether or not a regression is appropriate. As mentioned above a regression can be used for prediction, and as Carrol (1961, p. 48) stated, 'prediction is something you do after you have discovered the relationship between variables. But the prior measurement of relationship is important not only for prediction but also in its own right'.

INTERPRETATION OF THE PEARSON PRODUCT − MOMENT CORRELATION COEFFICIENT *R*

It is important to remember that r is a measure of any linear trend between two variables. The value of r will always lie between -1 and 1. If r is zero, then this indicates that there is no linear association between the variables. Note that there might be some nonlinear relationship but if r is zero then there is no consistent linear component to that relationship. If $r = 1$, then there is a perfect positive linear relationship between the variables, and all individuals sampled would lie exactly on the same straight line with a positive slope. If $0 < r < 1$ then there is a positive linear trend but sampled individuals would be scattered around this common trend line; the smaller the absolute value of r the less well the data can be characterized by a single linear relationship. If r is positive then an increase in the value of one variable would lead to our expectation that the other variable will also increase. An r value of -1 suggests a perfect negative relationship with any sampled individual always lying on the same linear trend line which will have a negative slope. If $-1 < r < 0$ then sampled individuals will be scattered around the line; again the smaller the absolute value of r the less well the data can be characterized by a single linear relationship.

Note that the magnitude of r gives no information about the gradient of the linear trend line; rather it gives a measure of how much scatter there is likely to be in a sample of individuals around that trend line. The value of r^2 is generally called the coefficient of determination. If $r^2 = 0.8$ then 80% of the variation between sampled individuals in their values of X can be explained by variation in their values of Y, and equivalently 80% of variation in Y can be explained by variation in X. The importance of r^2 in regression demonstrates that r is a useful measure even when regression is appropriate because a functional dependence between X and Y exists.

The value of r is independent of the units of measurement used to measure X and Y.

UNDERLYING ASSUMPTIONS REQUIRED FOR CALCULATION AND INTERPRETATION OF *R*

Traditionally, the underlying assumptions made in using r as a measure of association are that (1) the individuals in the sample are statistically independent of each other, and (2) the population from which the sample was drawn has a bivariate normal distribution for the two traits of interest (Whitlock & Schluter, 2009). The first of these assumptions is a fundamental of much statistical testing and does not require further discussion here. In fact, the second assumption is not absolutely vital and the correlation coefficient is informative about the degree of linear association between the two random quantities regardless of whether their joint distribution is normal. The characteristics mentioned above (i.e. $-1 \leq r \leq 1$, all points lie on a straight line if $|r| = 1$, and the smaller $|r|$ is the greater the amount of scatter around the trend line) are valid without the assumption of normality, or any other distributional assumption (Binder, 1959). Moreover, zero correlation can be an interesting null hypothesis without assuming bivariate normality, and can be tested with a permutation test which involves no assumption about a specific distribution (Pitman, 1937).

However, the calculated value of r from the sample is guaranteed to be the maximum likelihood estimate of the population